TERRAIN CLASSIFICATION USING LANDSAT THEMATIC MAPPER AND DIGITAL TOPOGRAPHIC DATA IN THE BURWASH UPLANDS, SOUTHWEST YUKON

JOAN ELIZABETH MOULTON

TERRAIN CLASSIFICATION USING LANDSAT THEMATIC MAPPER AND DIGITAL TOPOGRAPHIC DATA IN THE BURWASH UPLANDS, SOUTHWEST YUKON

by

©

Joan Elizabeth Moulton, B.Sc.(Hon.)

A thesis submitted to the School of Graduate
Studies in partial fulfillment of the
requirements for the degree of
Master of Science

Department of Geography
Memorial University of Newfoundland
29 January 1989

St.John's, Newfoundland

ISBN 0-315-50449-8

Abstract

Landsat Thematic Mapper (TM) images, if analysed properly, can provide land scientists with valuable terrain information. In high relief environments digital classification accuracies to date have been relatively low compared to those in less mountainous terrain. While the nature of the topographic effect on Landsat TM data is not fully understood, it is expected that low accuracy may be attributed, in part, to the lack of an appropriate expression of topography in the Landsat image data set. This study was designed to investigate the influence of various surface cover and topographic parameters on the spectral response measured by the TM sensor and show that a data set composed of topographic terrain descriptors can provide additional information which can be incorporated in terrain analysis of mountainous regions. A second objective was to investigate the improvement in TM terrain classification accuracy that could be achieved for a mountainous area in the Southwest Yukon if an ancillary topographic data set was incorporated in the analysis as a logical channel in a discriminant type classifier.

Correlation procedures were employed to systematically analyse the relationships between TM spectral response and the topographic component of terrain. Bivariate and multiple correlation coefficients were interpreted to show that landcover and topographic characteristics of the landscape are linked and that the parameters of both these components have an effect on TM data. Canonical correlation coefficients were interpreted to mean that the variance in the sensor data set was not fully explained by the variance in either the surface cover, topographic or combined data sets. This suggested that additional information may be contained in the topographic variables which is not contained in the sensor data and may be useful for classification in high relief terrain.

Two supervised classification schemes were used to investigate the improvement in terrain classification accuracy that was possible by incorporating topography. These classifications conform to the general principles of the

'landscape approach' and were based on nine biophysical classes studied in the field and in metric aerial photography. The first classification examined the statistical improvement in classification accuracy that was possible by augmenting spectral TM data with elevation, slope, aspect, relief, and percent vegetation cover measured at 672 pixels in the field. Discriminant functions were generated based on the TM data alone and integrated with the other terrain descriptors in several combinations. Classification accuracy was tested using 102 pixels which had not been used in the derivation of the functions. The results show that overall classification accuracy improved from about 64% when the TM data were used alone to 79% when elevation alone was added and up to 98% when the additional topographic field descriptors were used. Accuracy was 100% when the percent surface cover variables were included.

The second classification scheme examined the spatial impact of incorporating topography in the classification. This involved a maximum-likelihood classification and mapping of the entire study area using the TM data alone and, subsequently, the spectral plus topographic descriptors extracted from an interpolated digital elevation model (DEM) for all pixels in the study area. Mapping accuracy was 55.8% when the TM data were used alone and 77.6% when topography was incorporated. These results provide evidence that TM and topographic data sets derived from a DEM can be integrated in terrain classification to improve the accuracy of results in high relief environments.

Acknowledgements

Financial and logistical support was provided by Natural Sciences and Engineering Research Council research grants to Dr. S.E. Franklin, the Canada Centre for Remote Sensing Technology Enhancement Program and the computer facility at Newfoundland Oceans Research and Development Corporation (NORDCO) Limited through the Provincial Department of Development and Tourism. Additional support was provided by the School of Graduate Studies at Memorial University of Newfoundland in the form of University Fellowships and Teaching Assistantships from the Department of Geography. Support for the field component was provided by the Northern Scientific Training Program of the Department of Indian and Northern Affairs.

I would especially like to thank my supervisor, Dr.S.E. Franklin, for his advice, guidance, support and encouragement throughout my graduate program. Many thanks, to all those at NORDCO Limited, especially Mr. Randy Gillespie, and to Mr. Tim Perrott of CCRS for his instruction in using the Applied Resource Image Exploitation System (ARIES III). Special thanks to Mr. Derek Peddle for providing important DEM software and computer expertise. Thanks also to Mr. William Jones and Mr. Robert Rowsell for expert assistance in the field. The assistance of faculty in the Department of Geography and the staff at Computing Services, Memorial University is appreciated. Thank-you to Nicollette Sullivan and fellow graduate students Jacqueline Gallagher, Ngiap-Puoy Koh, and James Hooper for continued support. Lastly, I would like to thanks my parents who have always been a constant source of support and encouragement.

Table of Contents

List of Tables

List of Figures

Chapter 1
Introduction

1.1. Introduction

Accurate and up to date earth surface or *terrain* information is a common requirement of scientists in a wide range of disciplines since the nature of the surface affects the practical use of the land. Foresters, ecologists, engineers, geographers, and others concerned with land-related activities each use information about certain land attributes in daily decision making and planning. Information requirements include measurement of attributes which describe the surface cover types and morphometry. More specifically, the attributes may describe the vegetation type and composition at a site or the topography in the form of slope, aspect or elevation.

Traditionally, aerial photographs and ground surveys have been the main sources of terrain data. Aerial photographs are manually interpreted with the assistance of field knowledge to measure the associated surface attributes (Webster and Beckett, 1970). This approach is highly dependent on an experienced interpreter who is trained to recognize important terrain attributes for a specific application. Consequently, there are problems in acquiring terrain data associated with subjectivity, reliability, and repeatability.

Since 1972, Landsat satellites have orbited the earth and sensors on board the satellites have acquired *images* of the surface (Freden and Gordon, 1983). The two primary sensors on board the Landsat satellites are the Multispectral Scanner (MSS) and the Thematic Mapper (TM). The MSS was the initial sensor launched on the first of the Landsat series. Subsequently, it has been employed on all the Landsats in the series (five in total to date). On Landsats 4 and 5, the more advanced TM accompanied the MSS.

The Landsat images recorded by these sensors are digital representations of earth reflectance in different regions of the electromagnetic spectrum. All terrain features in an area sensed at a given point in time as well as the atmosphere contribute to the signal recorded at the satellite level (Robinove, 1979, 1981). The regions of the spectrum recorded by each Landsat TM are presented in Table 1-1. If analysed properly, these data may be converted into useful information that is valuable for resource managers and other land-related specialists.

Table 1-1: Landsat Thematic Mapper Bands

Band	Wavelength(μm)	Colour
1	0.45-0.52	Blue/Green
2	0.52-0.60	Green
3	0.63-0.64	Red
4	0.76-0.90	Near Infrared
5	1.55-1.75	Near-mid Infrared
6	10.40-12.50	Thermal
7	2.08-2.35	Middle Infrared

The operational use of Landsat data in terrain analysis has been demonstrated in a wide range of applications (Conner and Mooneyhan, 1985) Probably the most beneficial application in terms of resource management is *terrain classification*. This involves the delineation of earth surface regions that are similar according to certain attributes and may incorporate similarities in vegetation, soils, or geomorphometric attributes. When applied to Landsat multispectral data, classification involves the development of rules based on the spectral attributes of the surface or unique spectral signatures of different land cover types. The ability to measure specific attributes will depend on the degree of correlation between the attributes and the values recorded by Landsat. Subsequently the success of terrain classifications using Landsat data will depend on the nature of these relationships and the ability of the sensor to record the necessary information ie. the ability of the sensor data to act as suitable surrogates for the terrain attributes deemed necessary for the classification by the resource manager.

In relatively low relief regions, the terrain attributes which influence spectral response are primarily those related to land cover (Colwell, 1983). Different vegetation and soils attributes reflect varying amounts of energy throughout the electromagnetic spectrum and consequently have different spectral response curves. These differences in electromagnetic energy are recorded by Landsat sensors. The Landsat measurements can, therefore, be used as surrogates for mapping the distribution of vegetation types. Results of this type of analysis in areas of low relief have generally been successful (Kan and Weber, 1978; Mayer et al., 1979)

In regions of high relief, topographic variability, in addition to land cover attributes, is known to influence the data recorded by Landsat sensors (Colwell, 1983; Holben and Justice, 1981, 1980; Justice, 1978). Holben and Justice, 1981 defined this *topographic effect* as the variation in radiance from inclined surfaces compared to the radiance from a horizontal surface as a function of the orientation of the surface to the light source and sensor position. Numerous researchers have tested and quantified this effect in Landsat data: Holben and Justice (1980) showed that a range of fifty pixel values were associated with a single land cover type on a high solar elevation Landsat MSS image; Stohr and West (1985) and Dave and Bernstein (1982) provided evidence that variations in MSS data for a single cover type could be partially attributed to changes in the slope and orientation. Since the topographic effect can cause a wide range of pixel values to be recorded for a single land cover type, Siedel et al. (1982) suggested that if pixels for a given cover type overlap with values for other land cover types, pixels in the overlap regions may be incorrectly classified. Limited success in terrain classification of high relief regions suggests that topography acts as a source of inconsistency in the data and thus may be a source of error in terrain classification. In other words, in mountain environments, high topographic variability makes Landsat data less suitable as surrogate measures of terrain.

Early research in improving classification accuracy concentrated on developing an understanding of nature of the topographic effect on Landsat data. The main objectives focussed around determining the effects that various topographic-related parameters have on the remotely sensed data. This was necessary in order to determine if the spectral data could be employed to effectively monitor cover types. Knowledge of the environmental variables which influence spectral response will improve the ability to interpret cover type classifications as well as provide information which will help to determine the most effective way to combine variables in the classification procedure.

Numerous attempts have been made to remove or reduce this source of error by correcting the data for topographic effects. This approach is suitable for certain applications where the surface cover character is of primary importance. However, often in mountainous regions, the geomorphometric character of the terrain is important to the resource manager and consequently geomorphometric attributes may be necessary as classification criteria. For example, when classification extends beyond the simple separation of land in terms of a single terrain attribute such as vegetation, the approach is often referred to as an *Integrated* or *Biophysical* approach. *Landscape classes* or terrain classes are defined as regions with similar patterns of land form, vegetation, and soils. In this application, the removal of confounding topographic information in the Landsat data set is still an important operational problem in high relief terrain analysis, but a more direct approach is to consider the topographic information together with spectral information.

The basic idea is to improve classification accuracies in high relief terrain by the integration of two different data sets: (i) spectral response from satellites and (ii) topography derived from digital elevation models (DEM). This concept applies not only for Landsat spectral response; other types of imagery can benefit from the use of topography, for example, RADAR (Hinse et al., 1988) and Systeme Pour l'Observation de la Terre (SPOT) (Jones et al., 1988) imagery.

DEMs are similar to Landsat spectral images in that they are quantitative representations of the earth surface; however, each number in the model represents terrain elevations at known positions rather than spectral intensities (Burrough, 1986). Such models can be generated independently from ground survey (Brinker and Wolf, 1984), topographic maps (Collins, 1975), aerial photography (Crawley, 1974), or most recently, stereo space imagery (Cooper et al., 1985). The application of DEM data in digital terrain classification has been proposed and attempted for MSS data by Hutchinson (1978) Robinove (1981) and Franklin (1987); but no similar effort has been documented using the Landsat Thematic Mapper imagery in a subarctic environment, although several researchers (for example, Walsh, 1987) have pointed out the value of this approach for TM.

1.2. Statement of the Objectives

The main objective of this research is to determine whether or not a data set composed of topographic terrain descriptors can provide additional information and lead to improved classification results if integrated with spectral data acquired by Landsat TM in a high relief region.

The first step in achieving this objective is to determine statistically if in fact additional information is available from the topographic data set. This requires a systematic analysis of the relationships between spectral data acquired by Landsat TM and the topographic component of terrain in the region selected for this study. The following tasks are necessary:

(i) measure vegetation and topographic attributes at a random selection of sites within the study area,

(ii) extract the spectral response at each of these sites from the Landsat image containing the study area,

(iii) perform correlation analysis between the spectral data and ground variables measured at each site.

The next step is to determine the statistical improvement in landscape or terrain classification accuracy that can be achieved by integrating spectral and topographic data. The statistical analysis will be based on a discriminant function derived from (i) spectral data alone, (ii) topographic data alone, and (iii) both spectral and topographic data based on sites visited in the field. The difference if any between i, ii, and iii in terms of classification accuracy will provide evidence for the hypothesis that, in this region and in this application, spectral data must be analysed in conjunction with topography.

Finally, a spatial analysis of the integration of spectral and topographic data in the form of a map product is needed. The spatial analysis will be based on the entire study area and will involve:

(i) digitization of contours of topographic map of study area, interpolation of an elevation grid, and creation of a digital elevation model (DEM);

(ii) extraction of geomorphometric terrain attributes, elevation, slope, aspect, and relief from the DEM using available software;

(iii) resampling of topographic DEM data and spectral data to UTM coordinates;

(iv) performing supervised classification of the study area using spectral data alone, DEM data alone, and the integrated spectral and DEM data set using a maximum likelihood classifier;

(v) determining the differences between each classification in terms of the spatial effects of integrating the DEM data set and the practical use of the classifications.

1.3. Thesis Organization

This thesis is divided into seven chapters. The first chapter introduces the subject of terrain classification using data acquired by the Landsat series of satellites. Specifically, it describes the problem of obtaining accurate classifications in regions with high relief terrain.

The second chapter provides a review of previous studies that have (i) investigated the effects of topography on the data and (ii) attempted to improve accuracy rates in mountainous terrain.

Chapter three describes the methodology of this experiment including details of the acquisition and organization of ground data, spectral data, and topographic data employed in this research.

Chapter four and chapter five explain the statistical analysis performed on the data. In chapter four, the relationships between terrain attributes and the spectral data are investigated through various types of correlation analysis. In chapter five, statistical classifications and accuracy assessments are presented based on the spectral data alone, the topographic data alone, and the integrated data sets.

Chapter six contains a detailed description of the map production process based on the individual and integrated data sets. Maps are included to reveal the spatial distribution of the effects of topography on multispectral classification.

In the final chapter the conclusions and recommendations that arise from the research are discussed.

Chapter 2
Related Research

2.1. Introduction

Landsat data have been used for classification of terrain in many regions of the Earth. In high relief environments, however, classification accuracies using multispectral Landsat data have been poor and results been less than satisfactory. For example, Franklin and LeDrew (1984b) performed land cover classification in the Southwest Yukon using Landsat MSS data and achieved an accuracy level of only 58 percent; Fleming and Hoffer (1979) mapped forest types in the San Juan Mountains with 49 percent accuracy. In relatively flat terrain, corresponding accuracies for forest type mapping using MSS data were 91 percent (Mayer et al., 1979); landcover mapping in a lowland region of England and Wales using TM imagery was found to be 93 percent by Deane et al. (1985). These classification results suggest that low levels of accuracy obtained in digital classification of mountainous regions may be largely a result of the influence of topography.

Improving the accuracy of terrain classification using Landsat data has been an important subject in remote sensing research. General attempts to improve classification results have involved: (i) improvements in sensor design (Khorram et al., 1987), (ii) better radiometric correction of images prior to analysis (Ahern, 1985; Ahern et al., 1987), (iii) removing the effects of the atmosphere and topography through sophisticated radiometric calibration of satellite data (Robinove, 1982; Moulton, 1986), (iv) the development of more sophisticated classification algorithms, (v) more sophisticated processing methods for the data such as the extraction of texture (Franklin and Peddle, 1987; Haralick et al., 1973) or topographic information (Cooper et al., 1985; Wang et al., 1984) from digital images to be incorporated in the classification procedure, and (vi) the use of multitemporal and ancillary data sets (Franklin et al., 1987; Cibula and Nyquist, 1987; Gregory and Moore, 1986; and Satterwhite, 1984).

More specifically, attempts to improve classification accuracy which focussed particularly on high relief terrain required an understanding of the topographic variables that influence the data acquired by Landsat. Initial research focussed on developing this understanding. Once the topographic parameters which influenced the data were determined, image correction models were developed which could be employed to remove the effects of topography prior to classification. But in some applications, particularly integrated or landscape classification, topography was considered an important component that must in some way be incorporated into the classification process. The low levels of accuracy obtained in digital classification of high relief regions may be a result of the lack of an adequate description of the topographic component of the landscape. Rather than removing the effects of topography, some researchers felt that the integration of ancillary topographic information with the spectral response acquired by Landsat sensors would be a more suitable approach. Again, an understanding of the relationships between environmental variables and the spectral response was necessary to determine which variables were important as ancillary information upon which to discriminate the land cover classes of interest.

2.2. Nature of the Topographic Effect

Initial studies that investigated the nature of the relationship between topographic variables and Landsat data focussed on the use of Multispectral Scanner (MSS) data. For example, Justice (1978) showed that it was possible to determine the principal ground properties affecting MSS sensor response by correlating field measurements of ground properties to quantitative sensor data. Results in a Mediterranean study area indicated that the ground properties describing the vegetation composition at a site had the greatest effect on sensor data. The relationships between morphological variables, such as elevation, slope, and incidence, and sensor data were weak but statistically significant for the data analysed. In another study, Franklin and LeDrew (1984a) examined the relationship between spectral response patterns and surface geomorphological attributes for a high relief region in the Southwest Yukon. Geomorphometric variables in the form of elevation, slope, aspect, relief, and convexity were extracted from a digital elevation model and were correlated with Landsat MSS data. Results indicated that topography was an independent source of information which could be used with MSS data to improve classification of landscape units. Similarly, Walsh (1987) investigated the variability of MSS spectral response in relation to stand and site characteristics and concluded that DEMs should be included in the analysis of all spectral response patterns of mountainous regions.

For this research, data acquired by the more advanced TM sensor were available. The Thematic Mapper (TM) sensor acquires terrain spectral response data with improved spatial and radiometric resolution (Toll, 1985; Engel, 1983; Freden and Gordon, 1983; Chavez et al., 1983) compared to the data obtained by the Multispectral Scanner (MSS). While each MSS pixel represents a ground area of approximately 60 x 80 meters, the more recent TM sensor acquires data over areas 30x30 meters. The MSS sensor acquires data in four bands of the electromagnetic spectrum; the TM sensor records seven bands of data including a thermal band. The different spatial and spectral characteristics of MSS and TM data are suspected to show different or varying topographic effects for two main reasons: (i) topography affects different regions of the spectrum by varying amounts (Justice et al., 1981) and (ii) in complex environments slopes may vary considerably over relatively small regions, therefore, pixel values representing smaller sampling units may have a much higher degree of variability.

Karaska et al. (1986) investigated the impact of environmental variables on the spectral response of land cover recorded by the more advanced TM sensor. The spectral response of each of the seven TM channels was statistically tested against collected ground data on eleven environmental variables including slope, aspect, and surface roughness. Step-wise multiple regression analysis indicated that the percentage of trees and shrubs were most important in influencing the spectral response and that as the percentage of trees and shrubs increased, the more the effects of the other variables were obscured. Environmental variables related to geomorphology were found to have little effect on the data because very little change in elevation or topographic relief was present in the area investigated. A similar study by Hall-Konyves (1987) investigated the relationship between various topographic parameters and Landsat TM and MSS data in an area of gently undulating terrain. In a linear correlation analysis and analysis of variance, calculated incidence angle values and values of slope aspect and magnitude were integrated with MSS and TM data on a pixel by pixel basis. The results indicated that the relationship between topographic parameters and the Landsat data were weak for cultivated fields and forest areas. It was concluded that the topographic effect in such an area was of little importance.

Few studies have investigated the influence of topographic parameters on TM data in mountainous terrain where the effects of topography can be expected *a priori* to be much greater. This is in part because TM data have only been available since 1982 and most studies with TM data to date have been performed in relatively flat areas, for example, in agricultural applications. Furthermore, methods to successfully investigate the relationships between ground variables and Landsat data had to be established in relatively simple areas before attempting such an analysis in a more complex mountainous environment.

Based on knowledge that topographic parameters influence spectral response in high relief regions, two basic approaches have been taken in attempts to improve the classification of Landsat data in such areas. These have involved either (i) *elimination* or reduction of the topographic effect on the data using band ratioing or image correction techniques and (ii) *integration* of topographic information with spectral data to provide additional information upon which to discriminate terrain classes. These are discussed in the next two sections.

2.3. Reducing the Influence of Topography on Landsat Digital Classification

A very simple and straightforward technique used to reduce the topographic effect on Landsat data which requires no ancillary data is *band ratioing.* Band ratioing involves the creation of new channels of data by dividing each pixel value in one spectral band by the corresponding pixel value in another band (Richards, 1986; Bernstein, 1978). The effects of topography are assumed to be multiplicative and by ratioing the bands the multiplicative terms should cancel out (Woodcock, 1982; Holben and Justice, 1981). One major disadvantage of this approach is that it reduces the dimensionality of the data and often removes valuable information related to the brightness of pixels.

Holben and Justice (1981) examined band ratioing of data acquired by a ground based radiometer and showed that while this technique did not reduce the topographic effect on the data entirely, it did reduce the effect up to 83 percent for specific slopes and solar elevation angles. It was expected, however, that band ratioing of Landsat data would be less successful as a result of sensor calibration and quantization effects (Holben and Justice, 1981). Justice et al. (1981) examined the effect of the band ratioing technique on Landsat MSS data and found this expectation true; ratioing bands only slightly reduced the topographic effect.

After identifying the terrain parameters which affect multispectral response, researchers attempted to reduce the topographic effect inherent in MSS data by developing image correction models and applying correction algorithms to the data (Kawata et al., 1985; Teillet et al., 1982). Kawata et al. (1985) proposed a simple radiometric correction method which removes both atmospheric and topographic effects from remote sensing data and applied it to a mountainous site. Some success was achieved for a Landsat band 7 image; however, in regions where illumination was poor, the topographic effect was not removed. Justice et al. (1981) examined three models as methods of preprocessing Landsat MSS data for the topographic effect: a Lambertian image correction model; a modified Lambertian model; and a non-Lambertian model. The Lambertian model is based on the assumption that the surface being sensed scatters light equally in all directions and models radiance from the surface by the cosine of the incidence angle (the angle between the surface normal and the solar beam). This model was found to increase the effects of topography due to the "inapplicability of the Lambertian assumption to model the bidirectional reflectance characteristics of the woodland surface" (Justice et al., 1981, p.228). The modified Lambertian model produced higher variances than those found in the raw Landsat data. While the non-Lambertian model did decrease the topographic effect in this area, it must be evaluated in an area with a greater diversity of cover types before it can be applied in a more complex region. Smith et al. (1980) also evaluated the Lambertian assumption for Landsat MSS data and found that the Landsat response for ponderosa pine for incidence angles between 30 and 80 degrees and for exitance angles between 10 and 45 degrees does not follow the Lambertian law.

Cavayas (1987) examined the modelling and correction of the topographic effect on satellite image radiometry in a forestry context by using a bidirectional reflectance (BRF) model to correct two images acquired under different sun elevation and azimuth angles and comparing the reflectance estimated for each Landsat MSS band 7 on a pixel by pixel basis. A digital terrain model was used to derive slope and aspect parameters which were required as input to the BRF model. His results demonstrate that the analysis of multidate satellite images in conjunction with a digital terrain model can provide the means for a more thorough understanding of the topographic effect problem and permit the classification of forest covers with an accuracy comparable to or better than that of forest cover maps obtained by photo interpretation.

2.4. Data Integration

Other studies have attempted to improve terrain classification by incorporating ancillary data in digital classification (Peddle, 1987; Franklin et al., 1987, 1988; Richards, 1986; Shasby and Carneggie, 1986; Hutchinson, 1982; Strahler et al., 1980). Hutchinson (1982) described several ways in which ancillary data and Landsat data can be combined in the classification process. One method, *preclassification scene stratification* involves division of the study area into strata based on some criterion such as topographic data prior to the implementation of a classifier. In this way each stratum may be processed separately and it is possible to differentiate objects which are spectrally similar. Another method of integration involves *postclassification sorting* in which a large number of spectral classes are produced and then merged into groups which represent object classes. Problem spectral classes are assigned to the appropriate object classes using the ancillary topographic data set. A third method of integration, termed the *logical channel* approach by Strahler et al. (1978), involves increasing the number of observation channels during the classifier operations.

Woodcock et al. (1980) stratified a high relief region in Northern California into natural regions based on elevation and aspect prior to classifying the entire scene. Then, they used Landsat data and texture data to classify each natural region into height and density homogeneous forest classes. Although a quantitative evaluation of the accuracy of the final classification was not provided, qualitative assessment showed that the classification was similar to one produce by photointerpretation.

The *logical channel* approach was used by Strahler et al. (1978) and forest cover classification accuracies were improved by 27 percent when elevation was the additional channel. Similarly, Franklin and LeDrew (1984b) improved classification accuracies from 58 percent when the classification was based on spectral data alone to 87 percent when geomorphometric terrain descriptors were included as additional channels during classifier operations. Fleming and Hoffer (1979) improved classification accuracies by 19 percent by incorporating topographic data during the classifier operations.

Bonner et al. (1982) used the postclassification refinement technique and improved overall classification accuracies from 54 to 73 percent when elevation decision rules were developed for each class and pixels were reclassified with specific elevation breakpoints for each computer class.

Previous studies which used Landsat data in classification of a mountainous environment have all shown significant improvements in classification accuracy when topographic data were integrated at some stage in the classification process. These studies have all been performed on Landsat MSS data and, in each case, the method was based on an intimate knowledge of the relationships between terrain variables, such as slope, aspect, and elevation and the spectral response patterns. No similar classification improvements have been performed on the more recent TM data. This is partly because the relationship between topographic variables and Landsat TM data is not yet fully known.

2.5. Summary

Based on knowledge of the parameters that influence multispectral data, models have been developed and used to correct image data for topographic effects. Use of such models has obtained limited success in improving classification error rates with digital Landsat data in part as a result of the difficulty of accurately modelling topographic effects in high relief terrain. The nature of the relationship between topographic variables and MSS data in high relief environments has provided a rationale for the integration of spectral and ancillary topographic data in the classification process. Classification accuracies were improved considerably in all cases where topographic data was incorporated with Landsat MSS data either before, during or after the application of a classification algorithm to the data.

The relationship between topographic variables and Landsat TM data have been investigated only in regions of relatively low relief. Topographic effects have consequently been found to be weak or insignificant. The effect of various environmental variables on TM data has yet to be determined for a high relief region and is the first stage of this research. Further, since these relationships are not yet understood, classifications which incorporated both TM spectral data and topographic data have not been performed in a high relief environment where topography is an important component of the landscape. The second stage of this research will be to (i) develop an appropriate methodology for data integration based on the relationships identified in stage 1 and (ii) to perform integrated classification to determine the level of improvement that can be achieved if both spectral and topographic data sets are employed.

Chapter 3
Methodology

3.1. Introduction

In this chapter, the methodology and data employed in the Yukon study are described in two distinct sections. The first investigates the relationships between and among terrain variables and sensor variables. The methods employed here are modelled after those used by Franklin and LeDrew (1984a) and Justice (1978) in their studies of MSS data. The analysis employs correlation procedures which are discussed in full in chapter 4 (see also Thorndike, 1978 and Clarke, 1975). Analysis was performed using the Statistical Analysis System (SAS) (Helwig and Kathryn, 1979).

The second part of the methodology involves terrain classification which requires an understanding of the relationships between variables examined in section one. Two types of classification are performed. The first type is based on the generation of discriminant functions (Klecka, 1980) using SAS and is essentially an examination of the statistical improvement in classification accuracy that can be achieved by integrating data sets (see Franklin and LeDrew 1984b). The second classification procedure involves a spatial analysis and mapping of the complete study area using a Bayesian Maximum Likelihood Classification algorithm available on an ARIES image analysis system (DIPIX, 1987) . This classification examines the spatial impact of incorporating topography in classification.

Both classification procedures employ a *supervised* training approach. This approach requires that an operator select areas in the digital image which represent the terrain classes of interest to the image analysis or ultimate map user. It is essential that the operator has some knowledge about the study area either obtained through field investigation or through the analysis of aerial photographs, topographic maps, or any other data sources that are available. Unlike an *unsupervised* training approach where classes are defined based on the statistical structure of the digital data sets and which requires little operator input, known information can be input to the analysis. Further, since terrain classes are defined *a priori* to the application of the classification algorithms and based on known ground information when the supervised approach is adopted, the resulting classes will be of interest to the map user. If an unsupervised approach is used, this is not always the case; statistical divisions or clusters in the data set do not always correspond exactly with the classes one wishes to discriminate. Supervised classification is considered to be a more powerful test of terrain analysis applications and was consequently employed in this research.

The classification scheme developed for both the discriminant and maximum likelihood analysis is an *Integrated* or *Landscape* one in which classes are defined in terms of similarities in surface cover and morphometry. Concepts involved in this approach are discussed by Christian (1958) . Bastedo et al. (1984) describe a similar approach as an *ecological* approach to resource surveys (see also: Christian and Stewart, 1968; Mabbutt, 1968; Hutchinson, 1978; Robinove, 1979, 1981). This classification scheme is particularly suitable for the Yukon environment where few detailed surveys have been carried out. In such areas a fast and accurate reconnaissance survey as provided by the digital classifications can be used to select areas with potential for development of a particular land-use (Bastedo and Theberge, 1983) . More detailed surveys can then be performed only in selected areas. In high relief environments, both surface cover and topographic characteristics influence the utility of the land. The integrated or biophysical classification provides information on both of these components and further shows interrelationships between them throughout the area. This is valuable information to the land manager who must consider all aspects of the terrain when determining the optimal potential and utility for a region. It should be kept in mind, however, that the incorporation of topographic information in terrain analysis of high relief environments may also be important for improving results in more specific classification applications, for example, mapping geological, geomorphological, or glacial landscape classes and that data integration is not limited to classifications based on the landscape approach.

The results of the discriminant classifications are presented in chapter 5 as *classification accuracy* and *interpretation accuracy* after Franklin (1987, p.63). Classification accuracy is a measure of the ability of the discriminant functions to separate the the pixels used to generate the functions. Interpretation accuracy is calculated using an independent test sample and measures the capability of the discriminant functions to separate the terrain classes of interest in the study area. In this study, two sets of training and test pixels (both extracted randomly from the field data) are used to generate and test the discriminant functions. Two-groups were used to ensure that the pixels selected adequately represented the data set as a whole.

The results of the spatial analysis are presented in chapter 6. For discussion purposes, assessments of the maximum-likelihood classifications are documented as *mapping accuracy* where mapping accuracy is a measure of the agreement between classes identified on the digital map product and those identified in the field. In this analysis, mapping accuracy is calculated using 774 pixels known from ground survey.

3.2. Study Area

The study area, located in the Kluane Ranges of the Southwest Yukon (Figure 3-1), was selected for several reasons: (i) cloud free TM data and aerial photography were available; (ii) the area contains high variability conditions with simple landscape components; (iii) it is close to an area of previous research from which this study can gain experience; and (iv) it is easily accessible by the Alaska Highway which runs through the Northeast section and a cart track that permits vehicle access to the interior.

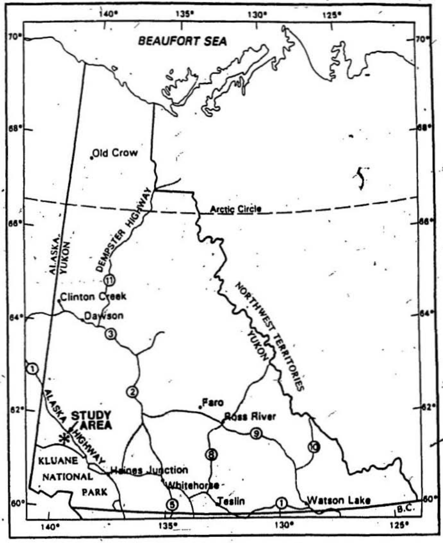


Figure 3-1: Study Area Location - Southwest Yukon Territory

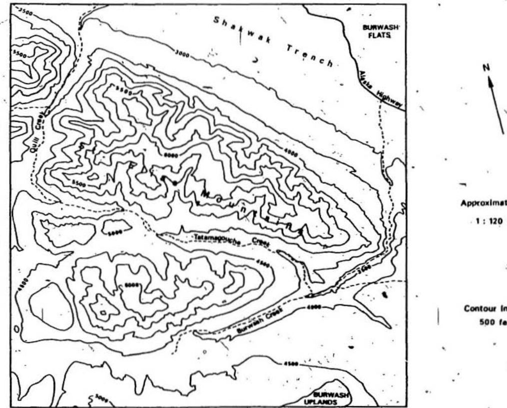
The study site (Figure 3-2) covers an area of approximately 250 square kilometres. Located in the Kluane Game Sanctuary, North of Kluane National Park, the area is of considerable environmental importance. Bounded to the Southwest by the St. Elias Mountains and to the Northeast by the Shakwak Trench, it is situated between two major fault systems, the Duke River and Denali, and is characterized by Carboniferous, Permian, and Triassic volcanic and sedimentary rocks (Theberge, 1980). The range of relief is greater than 1250 metres with a minimum elevation of 750 metres and a maximum greater than 2000 metres above sea level.

During a previous field season, it was noted that a considerable range of vegetation communities was present in the area which was intricately linked to topography and landform. While the topography of the area is complex, composed of variable slopes, aspects, and relief, the ecology is relatively simple, thus simplifying the identification of terrain units that are consistent in terms of landform, vegetation, and soils.

3.3. Data Acquisition

Two distinct groups of data were required for the study. The first group contains information for only a random selection of sites in the study area and includes: (i) TM data consisting of a value for each TM band at each site, and (ii) ground data derived directly from field measurements. The ground data consists of: (i) topographic data and (ii) surface cover in terms of a percentage of complete cover. All of these data were used to investigate the effect of various terrain properties on reflectance data recorded by TM, and to investigate the improvement in terrain classification accuracy that can be achieved by integrating TM sensor and topographic data.

The second group of data sets is comprised of (i) TM data for the entire study area and (ii) topographic data for each pixel in the study area derived from a digital elevation model. These two data sets were used to investigate the spatial effects of incorporating ancillary topographic data in terrain classification of the



Approximate Scale
1 : 120 000

Contour Interval
500 feet

Figure 3-2: Southwest Yukon Study Area complete study area. A diagram illustrating the various data sets and the procedures in which they are employed is presented in figure 3-3.

Spectral Data:

A computer compatible tape containing Landsat TM data of the study area was obtained from the Canada Centre for Remote Sensing (CCRS). The image was acquired 31 July, 1985 with a sun elevation of 44 degrees and azimuth 150. A subscene of 550 x 550 pixels which represents the study area was extracted from the image tapes using an ARIES III system at NORDCO Limited. A colour composite (bands 5, 4 and 3) showing most of the study area is presented in Figure 3-4. In actual fact, the image area is square; however, as a result of photographic reproduction, a strip at the top and bottom of the image area is not shown. Refer to Figure 3-2 for the complete image area.

The spectral data were used in two ways: (i) as a complete set, i.e. each pixel in the study area contained 7 values corresponding to the seven TM bands and (ii) a subset was extracted whereby only values which corresponded to the pixels in the field sites were contained in the data set. These values were extracted from the complete set by determining the line and pixel value of the center pixel of each site visited in the field. This was accomplished using a task on the ARIES III system which registers Landsat images to UTM coordinates. After the line and pixel coordinates were known for the center pixel of each site, a program, OUTWDW.FOR (see Appendix A) was written to extract the pixel with the given line and pixel coordinates and the surrounding eight neighbours for each site.

Ground Data:

Field data collection at specific sites was required to obtain information regarding the ground variables which may influence spectral response. A random selection of sites was necessary to ensure that an unbiased representation of the study area was obtained. Justice (1978) discussed several types of random sampling schemes. A purely random sample would have been difficult to execute in practice and it would have taken too long to gather the necessary field

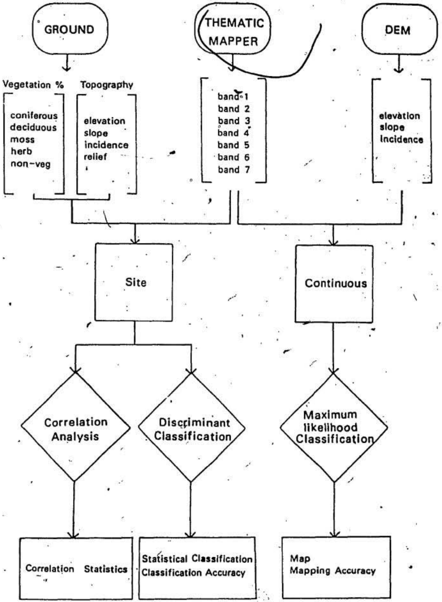


Figure 3-3: Data Sets Used in the Research

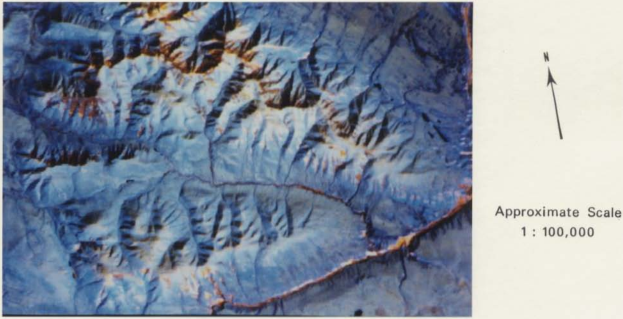


Figure 3-4: Landsat TM Colour Composite of Yukon Study Area - Bands 5,4,3

observations. A systematic strategy was adopted so sites could be visited in an orderly fashion. A disadvantage with a purely systematic sample in which sample sites are selected at regularly spaced grid intervals is that the regularity of the sampling may coincide with regularities in the terrain (as caused by aligned cuestas and strike vales) (Townshend and Justice, 1981, p.42). Consequently an unaligned systematic sample was adopted. Instead of selecting a random sample within each square of a constant grid, however, the X or Y coordinate was the constant element. A random series of X (easting) and Y (northing) coordinates was generated and each X was paired with each Y to generate a series of X/Y coordinate pairs. This modification simplified both generation of the random sites and location of the sites in the field. At the same time, it preserved the randomness required for future statistical analysis.

A sample site of larger than one pixel was required to permit accurate ground location in terms of Universal Transverse Mercator (UTM) coordinates (Justice and Townshend, 1981). Therefore, an areal sample was taken at each site which comprised a 3 x 3 pixel window and represented 8100 square metres on the ground. This was within the guidelines suggested by Justice and Townshend (1981) regarding minimum sampling unit for MSS data. Ground data were collected at 100 sites for a total of 900 pixels. This sample was believed to adequately represent the variability in the study area according to Hammond and McCullagh (1980).

The ground characteristics recorded were chosen to quantitatively describe the morphometry and surface cover at each site. For some areas, it may be argued that adequate surface information could be obtained through the interpretation of large scale aerial photography and topographic map sheets. In a complex environment, particularly the study area for this research, the high variability of terrain made field measurement of ground properties essential if the exact nature of the effect of topography was to be determined.

The surface cover variables were recorded at each site by an interdisciplinary field team in the summer of 1987. They included the percent cover of coniferous vegetation, deciduous vegetation, moss, herb, and non-vegetated cover and were measured in terms of percent coverage at each site as viewed from directly above the site. Percentages were estimated in 5% increments, for example, 5%, 10%, or 15% coverage and so on.

The topographic variables recorded at each site were selected with reference to the general system of geomorphometry described by Evans (1972) and used in conjunction with MSS data by Franklin (1987) to discriminate parcels of land from adjacent terrain in a study area in the Ruby Range, Southwest Yukon. Evans described general geomorphometry as the field of measurement and analysis of those characteristics of landform that are applicable to any continuous rough surface. The landform characteristics include elevation, slope, aspect, relief, and convexity.

Elevation is the height above sea-level of the site. In this study, elevation for a particular site was read from a 1:50,000 scale topographic map of the region.

Slope is the rate of change of altitude with distance and is calculated as the first (vertical) derivative of elevation. In this study, slopes were measured at each site from the center of the middle pixel to the center of each surrounding pixel. Based on these slopes an average slope plane was then calculated for the site (Appendix A). Each pixel was then given the average slope value in subsequent analysis.

Slope has a directional component known as aspect which is the first horizontal derivative of elevation. Aspect measured by Evans (1972) in degrees as the direction the ground faces, unfortunately , is not a metric; for example, 10 degrees is closer to 350 degrees than 40 degrees. As a result an alternative way of expressing aspect was required for this analysis. The approach taken in this study was adopted from Justice (1978) who computed incidence values as a function of aspect, slope, solar elevation, and solar azimuth. In this form, aspect could be incorporated in the necessary quantitative statistical analysis. The following formula was used to compute the incidence values:

$$inc = cos(\alpha) + sin(\alpha)\ cot(\beta)\ cos(\theta)$$

where α is slope, β is sun elevation, and θ is the difference between terrain aspect on a compass and the solar azimuth.

Evans (1972) also included convexity in his general system of geomorphometry. He defined convexity as the rate of change of slope relative to aspect and calculated it as the second derivative of elevation plus the first derivative of slope. Convexity was not incorporated in the research because it was believed that convexity measures would provide little addition information over an area 30 meters square which represents the size of each TM pixel. Franklin (1984b) used convexity measures in similar analysis based on MSS data which is comprised of larger size pixels (60x80) meters and found only slight improvements in analysis results when convexity was employed.

Digital Elevation Model Data:

Ground data collection made topographic data available only for a selection of sites in the study area. If the entire study region was to be mapped it would be necessary to acquire topographic data for every pixel. Therefore, a Digital Elevation Model (DEM) of the study was created and the necessary geomorphometric terrain variables were extracted from it for the entire area.

Creation of a digital elevation model of the area involved two main stages: (i) manual digitizing the contours of a recent 1:50,000 scale National Topographic Series (NTS) sheet of the area and (ii) creation of a regularly spaced grid of elevations by manipulating the digitized data. Digitization was performed using a Gentian Hi-State precision coordinate digitizer. A hand held cursor was used to record the location of X-Y coordinates while Z values were entered from the keyboard by the operator. Points were digitized along contours at the discretion of the operator. A general rule of two points times the distance between contour intervals was followed in most cases except in areas where contours were widely spaced and a greater number of points was required. At elevations less than 5000 feet, every contour was digitized; 5000 feet and above, every second contour was digitized depending on the closeness of contour spacing. In areas where contours were extremely close, the degree of operator error would reduce any improvement that could be achieved by digitizing additional contours.

Creation of the elevation grid from the digitized contours was done using the Surface II Graphics System (Sampson, 1978). This involved calculating elevations at grid nodes superimposed over the digitized data using a two phase local fit algorithm. In the first phase, a weighted trend surface was fit to each point based on an average projected slopes calculated for n nearest neighbours (Peddle, 1987) where n is equal to eight. In the second phase, a distance weighted average slope is calculated for each grid point using the trend surface equations developed in phase one. A grid size of 550 X 550 grid points was selected to correspond to the TM sub-image of the study area. The resulting DEM is presented in Figure 3-5.

Several sources of error are inherent in this approach and must be identified. An obvious source of human error results from the manual technique used to digitize contours and the subjectivity involved in selecting points to record along a contour line. Extreme care and checking procedures were adapted in order to minimize the human error source. An important systematic error exists in areas that are particularly flat. In such areas there are large gaps between contour lines. These areas often lack sufficient data input to interpolate values for the extremely fine elevation grid required here. In order to minimize problems that result from limited data, additional contours were interpolated by the operator between widely spaced contours and input to the digitized data set. While this did eliminate the problem of missing data values, it resulted in an artificial component in the model, for example, the step-like appearance in Figure 3-5.

Figure 3-5: DEM of Southwest Yukon Study Area

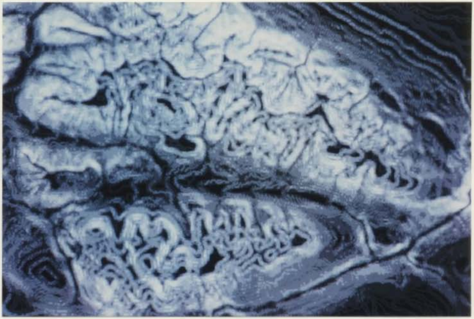
The artificial steps can be reduced to a certain degree by running a smoothing algorithm through the model (Hall-Konyves, 1987). The algorithm employed on the DEM for the Yukon study area used a filter and a cubic convolution resampler to average values within a 5x5 window successively throughout the image. The resulting smoothed DEM is shown in figure 3-6.

Software is widely available to extract geomorphometric terrain descriptors in the form of elevation, slope, incidence, relief, and convexity from DEMs (Collins and Moon, 1981; Franklin and Peddle, 1987) and to produce separate images which can be registered to the spectral data acquired by Landsat. The calculation of these variables is discussed in detail by Peddle (1987); Frew (1984) discusses image registration procedures in detail.

In this analysis, the software developed by Peddle (1987) was employed to extract raster slope, and incidence (a function of aspect) images from the smoothed Yukon DEM (Figure 3-6). These images are represented in Figures 3-7 and 3-8, respectively. On the slope image, light tones represent areas with steep slopes and dark tones represent relatively flat areas. Light tones on the incidence image represent areas with high incidence values which are calculated as a function of slope, aspect, solar elevation, and solar azimuth.

It is obvious from visual observation of Figures 3-7 and 3-8 that artifacts have been introduced into the slope and incidence images (for example, the linear features on Figure 3-7 which appear to follow the digitized contour lines and the step-like patterns on the relatively flat areas of Figure 3-8). These artifacts are a result of the systematic errors identified in the creation of the DEM. While steps were taken to minimize these problems as described previously, they could not be fully eliminated. The slope and incidence images that are provided were considered to be the best approximations of slope and incidence that could be acquired given the limitations of the manual technique for creating the DEM.

Figure 3-6: Smoothed DEM of Southwest Yukon Study Area



N

Approximate Scale
1 : 100,000

Figure 3-7: Slope Image of the Yukon Study Area

N

Approximate Scale
1 : 100,000

Figure 3-8: Incidence Image of the Yukon Study Area

3.4. Summary

The study area in the Southwest Yukon is characterized by three data sets: (i) spectral TM data for seven bands, (ii) ground data acquired through field survey, and (iii) data extracted from a DEM. Site data consists of spectral data and ground data for 774 pixels visited during the field season. Each of these pixels has a set a values which correspond to each TM band, elevation, slope, incidence, relief, and the percent cover of the various vegetation types. Continuous data for the whole study area consists of TM values and topographic data (elevation, slope, and incidence) derived from the DEM. In the next two chapters, various subsets of these data sets are employed in the analysis in order to determine if an integrated data set composed of digital TM and topographic data could be employed to improve terrain classification in a high relief environment and the level of improvement that could be achieved.

Chapter 4

Relationships Among Ground Variables and TM Spectral Response

4.1. Introduction

An understanding of the statistical relationships among ground variables and Landsat TM bands is necessary if TM data are to be used effectively for terrain analysis and classification. For example, an understanding of the nature of these relationships can provide insight into how certain terrain properties may be used to improve description and/or discrimination in the classification process. Statistical relationships between variables strongly influence classifier operations. Weak or moderate correlations between terrain properties and TM data may suggest that the associated terrain variables contain new information and if included as ancillary information in the classification process, may improve discriminating power; strong correlations suggest data redundancy; zero correlations suggest complete independence. Other reasons for examining the statistical relationships between the ground variables and sensor data are that a rigorous analysis of the environmental factors that are capable of significantly altering spectral response can be used to guide field work (Walsh, 1987) or generate new variables which reduce redundancies and contain aspects of both data sets (Franklin, 1987).

In this chapter, the results of correlation analysis performed among various ground variables and digital data acquired by the Landsat Thematic Mapper sensor are presented. Three types of linear models are studied in order to fully examine relationships between variables: (i) Bivariate analysis, which describes the relationships between each individual variable; (ii)

Multivariate/regression analysis, which identifies the relationships between sets of ground variables and each TM band; and (iii) Canonical correlation analysis, which reveals relationships between sets of ground variables and the set of seven TM bands.

The various statistical procedures were performed on a computer system for data analysis called SAS (Statistical Analysis System) (SAS Institute Inc., 1985). Three sets of data were input to SAS statistical procedures: (i) surface cover data which includes the percent cover of the different vegetation variables and the percent non-vegetated cover recorded for the nine pixels associated with each site, (ii) topographic data composed of the four topographic variables (elevation, slope, incidence, and relief) derived from measurements at each site, and (iii) sensor data consisting of the seven Landsat values recorded by the Thematic Mapper for the pixels corresponding to each site. Each data set consists of 774 data values for each variable. Descriptive statistics for each variable are presented in Tables 4-1.

4.2. Bivariate Correlation

Bivariate Correlation is a technique which is used to investigate the relationship between two variables (Thorndike, 1978). In the first stage of this analysis, bivariate correlation is performed among the various terrain properties recorded at each site and Landsat TM sensor bands. This is an attempt to understand the terrain properties or attributes that are capable of contributing significant variance to TM data. This study attempts to document these relationships for a particular area in the Yukon and gain insight into the factors affecting remote sensing image analysis in that area.

Results:

The bivariate relationships between each TM band and the topographic and land cover parameters measured at each site are presented in Table 4-2. Correlation coefficients (r) were calculated using the Pearson Product Moment Correlation statistic. The square of the correlation coefficient, known as the

Table 4-1: Descriptive Statistics for Ground and Sensor Variables

Variable (unit)	Mean	Std.Dev	Minimum	Maximum	Std.Error	Sum	Variance	C.V.
Band 1 (DN)	70.27	7.10	59.00	111.00	0.26	54390.00	50.40	10.10
Band 2 (DN)	29.09	4.87	19.00	52.00	0.18	22514.00	23.69	16.73
Band 3 (DN)	28.62	7.06	16.00	60.00	0.25	22153.00	49.90	24.68
Band 4 (DN)	63.41	16.94	19.00	121.00	0.61	49076.00	286.88	26.71
Band 5 (DN)	76.01	21.42	23.00	133.00	0.77	58833.00	458.95	28.184
Band 6 (DN)	122.20	6.41	104.00	138.00	0.23	94584.00	41.10	5.25
Band 7 (DN)	28.26	9.32	8.00	74.00	0.34	21876.00	86.82	32.97
elevation (m)	1244.44	317.48	794.00	1865.00	11.41	963193.00	100791.25	25.51
slope (deg)	16.50	13.12	0.00	51.00	0.47	12771.00	172.22	79.53
incidence (deg)	0.95	0.19	0.35	1.23	0.01	731.75	0.04	19.84
relief (m)	30.00	14.93	0.00	58.99	0.54	23215.23	222.99	49.79
coniferous (%)	20.99	24.62	0.00	80.00	0.89	16245.00	605.91	117.28
deciduous (%)	25.35	24.04	0.00	80.00	0.86	19620.00	577.95	94.84
non-veg (%)	8.55	21.21	0.00	100.00	0.76	6615.00	449.92	248.19
herb (%)	31.40	24.24	0.00	95.00	0.87	24300.00	587.77	77.22
moss (%)	13.72	13.31	0.00	60.00	0.48	10620.00	177.08	96.99

coefficient of determination, is a direct measure of the proportion of the variance explained by the linear correlation. A T-test is used to determine whether or not each correlation is significant. Correlation coefficients are not considered to be significant for probability values greater than .01.

Results of correlations between the topographic variables recorded at a site suggest that elevation and slope are the most strongly related with a correlation coefficient of 0.52 which is moderate. This relationship satisfies the general rule of increasing slope angle with increasing elevation in areas which have been modified by glacial erosion (Franklin and LeDrew, 1984a). Other significant correlations were identified between slope and relief (0.20). Significant correlations between these two variables were also identified by Franklin and LeDrew (1984a) who suggested that this may be a function of the possibility of high relief values in areas of high slope even if the slopes are smooth when relief is measured as the variance in elevation. Increasing slope with increasing elevation may partially explain a significant correlation between elevation and relief. Slope is the only variable with which incidence is significantly correlated. A negative correlation coefficient (-0.34) suggests that steeper slopes have low incidence values and thus tend to face in the direction opposite direct solar illumination when the satellite image was acquired.

Significant correlations between elevation and percentage coniferous (-0.73), herbaceous (0.55) and non-vegetated (0.28) surfaces illustrate a strong altitudinal control on vegetation as perceived during field investigation. A negative correlation between coniferous cover and elevation is explained by the fact that spruce trees occur predominantly in the valleys and on alluvial plains at relatively low elevations in the study area. Moderate positive correlation between herbaceous cover and elevation is due to the increase in herbaceous cover at higher elevations where tree growth is constrained. Significant correlation between non-vegetated surface cover and elevation is in part a function of the moderate positive correlation with slope. This is explained by the tendency for areas which lack vegetation to occur on steep slopes which, as expressed in the

Table 4-2: Bivariate Relationships

Variable	Band 1	Band 2	Band 3	Band 4	Band 5	Band 6	Band 7
Band 1	1.00						
Band 2	0.94	1.00					
Band 3	0.95	0.97	1.00				
Band 4	0.33	0.50	0.42	1.00			
Band 5	0.59	0.71	0.69	0.80	1.00		
Band 6	0.47	0.53	0.50	0.50	0.64	1.00	
Band 7	0.78	0.84	0.84	0.56	0.90	0.64	1.00
elevation	0.40	0.50	0.53	0.49	0.58	*	0.52
slope	0.16	0.21	0.20	0.20	0.17	-0.14	0.21
incidence	0.39	0.42	0.39	0.36	0.37	0.50	0.39
relief	0.33	0.33	0.34	0.12	0.24	0.19	0.31
coniferous	-0.29	-0.35	-0.36	-0.35	-0.40	*	-0.36
deciduous	*	*	*	0.24	*	0.10	*
non-veg	0.28	0.25	0.30	*	0.21	*	0.37
herb	0.14	0.22	0.23	0.31	0.32	*	0.23
moss	-0.29	-0.32	-0.28	-0.27	-0.30	-0.38	-0.31

* - denotes correlation not significant at 0.01 level of confidence

Table 4-2, continued

Variable	elevation	slope	incidence	relief
Band 1				
Band 2				
Band 3				
Band 4				
Band 5				
Band 6				
Band 7				
elevation	1.00			
slope	0.52	1.00		
incidence	*	-0.34	1.00	
relief	0.24	0.20	*	1.00
coniferous	-0.73	-0.47	0.10	-0.16
deciduous	*	*	0.12	*
non-veg	0.28	0.41	-0.09	0.16
herb	0.55	0.13	*	0.15
moss	*	*	-0.40	*

* - denotes correlation not significant at 0.01 level of confidence

discussion of topographic intercorrelations, are more likely to occur at higher elevations. Coniferous cover is also moderately correlated with slope (-0.47). This is a function of the general distribution of coniferous trees at low elevations on relatively gentle slopes. A negative moderate correlation (-0.40) between incidence value and percentage moss cover recorded at a site suggests that proportions of moss are higher on slopes with low incidence values. During the field investigation, it was noted that northward facing slopes had visibly larger percentages of moss vegetation. This is explained by the hardy characteristics of moss requiring generally less amounts of insolation than other herbaceous types of vegetation which are common at similar altitudes in the study area.

All significant relationships between percentage vegetation variables and TM spectral response are weak with the exception of percent coniferous cover with Band 5 which is moderate (-0.40). Correlations involving the coniferous variable are negative with all spectral bands and are generally stronger than the other vegetation variables ranging from 0.29 to 0.40. High proportions of coniferous vegetation are associated with low spectral response due to the high absorption characteristics of black spruce which is the dominate coniferous species in the area. Percentage deciduous cover is significantly correlated only with Band 4 of the spectral bands and the correlation coefficient is weak (0.23). This is probably due to the wide range of deciduous species that occur in the area. These species include poplar, willow, alder, and dwarf birch all of which have varying spectral characteristics. Similar results were achieved by Justice, 1978 between percentage deciduous vegetation and MSS data.

The correlations between the herbaceous and moss vegetation variables and the spectral variables are all significant, generally ranging from 0.20 to 0.31 with large proportions of herbaceous cover associated with high spectral response values and large proportions of moss associated with relatively low spectral response. These correlations are explained by the absorption characteristics of moss and the dominating reflection characteristics of herbaceous cover in all spectral bands. All correlations between the proportion of non-vegetated cover variable and spectral bands are weak with the exception of Band 4 which is insignificant.

A significant correlation was identified between the vegetation variable representing percent moss cover and the thermal channel acquired by the Thematic Mapper (Band 6). A negative correlation coefficient of -0.38 may be a function of the thermal properties of moss but is largely determined by the distribution of this vegetation type predominantly on northward slopes with low incidence values. This was noted in the field survey and suggested previously in the bivariate correlation between incidence and the percent moss vegetation cover variable. No other correlations between vegetation and Band 6 were significant with the exception of deciduous cover which was very weak (0.10).

Significant correlations exist between topographic variables and TM spectral response. In particular, elevation is moderately correlated with all spectral bands. Correlation coefficients range from 0.40 to 0.58. This may be explained largely as a consequence of the significant correlations between elevation and vegetation cover at a site and especially as a result of the strong correlation between elevation and percentage coniferous cover as discussed in a previous section. The strongest correlation exists between elevation and Band 5 which is consistent with results obtained by Franklin and LeDrew (1984a) with MSS Band 7 which is an equivalent infrared wavelength. The fact that correlations between topographic variables and spectral response are significantly higher than those between vegetation cover and spectral response suggest that the influence of topography on TM spectral response is, in part, independent of surface cover. This is further demonstrated by considering the correlations between incidence and TM spectral bands. Incidence is consistently weak-moderately correlated with spectral response with coefficients ranging from 0.36 to 0.42. This may be explained as a function of direct solar illumination on southward facing slopes resulting in higher spectral reflectance from the surface but is also a function of the vegetation cover with relatively high reflectance characteristics. Slope and relief are weakly and positively correlated with all spectral TM bands.

A moderate correlation (0.50) exists between the thermal TM band and incidence. This is a result of direct solar illumination on southward facing slopes as well as the thermal properties of vegetation species that occur on these slopes in contrast to indirect illumination and vegetation on opposite northward slopes.

Discussion:

The correlation among topographic variables recorded at each site indicate the degree to which each variable contributes to the overall topographic character of the region. All topographic variables are significantly correlated to at least one of the other topographic variables. Since the correlations are generally weak and at best moderate, this may support the hypothesis that each not only contains information, but that each contains 'unique' information. We have the important work of Evans, 1980 to indicate the uniqueness of each of these variables.

Significant relationships between the topographic variables and surface cover variables further support the importance of topographic variables in explaining variance in land cover. For example, a direct link between the surface cover and topography, could suggest an altitudinal control on vegetation. This information is important in that it supports an approach to terrain analysis and classification that incorporates both surface cover and morphometry, namely the integrated or biophysical approach discussed in the Introduction to this thesis.

The fact that significant relationships exist between percent surface cover variables and TM response indicates that there is information related to surface cover contained in the TM data. This is because the percent cover of a given surface cover type is in part an indicator of the type of vegetation. For example, if the percent cover is 100%, the vegetation described by this measurement cannot be black spruce because black spruce does not occur at 100% coverage. On the other hand, it may represent the percentage herbaceous cover in a open meadow field. Significant relationships between percent cover and MSS spectral response were also identified by Justice (1978).

Although the relationships between percent cover are significant, they are weak because percent cover is only a partial indicator of vegetation type. Consequently, other information may contribute to the TM response and may be necessary to record during field investigations for example, soils and topographic characteristics. Only then can the reflectance values be fully explained and interpretations made on how spectral response can be used to identify different surface cover types or different landscape types that contain vegetation characteristics as components.

These relationships indicate that topography has a direct influence on the remotely sensed TM image data. Further investigation is required in order to make firm conclusions as to whether or not the topographic element contained in the TM data set is a source of noise and confusion or if can be used to discriminant topographic characteristics within landscape classes.

Summary:

The preceding bivariate correlation analysis revealed that: there are significant relationships between surface cover variables and topographic variables in this area of the Yukon. This can be interpreted as evidence that an analysis approach that considers both topographic and surface cover characteristics is appropriate for this high relief environment, namely, an integrated approach. Furthermore, this information can improve future field work in high relief terrain because it provides evidence that topographic variables must be included in the field analysis. It is also shown that each topographic variable measured (elevation, slope, incidence, and relief) is unique since in no case does the variance in one topographic variable fully explain the variance in another topographic variable. And finally, the bivariate correlation coefficients revealed that each topographic variable contains variance that is unique from the sensor variables. Relationships between each of the sensor variables and each of the topographic variables are weak to moderate at best.

While this analysis indicates the uniqueness of individual variables (leading to the conclusion that each individual variable should be incorporated in terrain analysis) these correlations should be interpreted cautiously. For example, the analysis does not indicate how topographic variables jointly contribute to spectral response. Since terrain analysis such as classification considers numerous variables simultaneously in a discrimination process, the question of how the set of topographic variables jointly contribute to spectral response must be investigated. This is done in the next section using multiple regression and correlation.

4.3. Multiple Regression Analysis

Regression analysis is a techniques used for examining the relationships between one variable and a combination of two or more other variables that are considered simultaneously (Thorndike, 1978). In general, this approach permits an analysis of the relative impact of each input variable which can be interpreted in physical terms as *usefulness* (Walsh, 1987); for example, in a classifier. In this research, regression analysis was performed in order to examine the relationships between each TM band and the set of (i) surface cover variables, (ii) the set of topographic variables, and (iii) the combined set of surface cover and topographic variables (all ground variables). In many ways this analysis can be seen to complement and confirm the earlier bivariate analysis and the canonical analysis of the entire data sets discussed in the following section. Here we discuss only the complete model using all available descriptors to the importance of variability in topographic and land cover conditions affecting the recorded TM brightness values.

A complete understanding of the influence of ground variables on each individual TM band is important because in some situations not all TM bands may be available or necessary for the analysis. In this case, knowledge of these relationships may indicate the optimal selection or combination of bands. Many previous investigations (for example, Justice, 1978) used regression techniques to examine relationships between ground variables and MSS spectral response. This analysis will provide a basis for comparison of relationships in this study with those in other environments.

Results:

Results of the regression analysis are presented in Table 4-3(a-c).

Correlation measures were calculated using the method of least squares to fit general linear models. Correlations are expressed in terms of the multiple coefficient of determination (r^2) which describes the amount of variance in the dependent variable (TM band) explained by the independent set of variables. F-tests determined that the overall correlations were significant at the .01 level of confidence in each case. T-values identify whether or not the contribution from each independent variable is significant.

Results suggest weak relationships between the surface cover variables when considered as a group and the sensor data for each TM band Table 4-3(a). For example, the r^2 ranges between 0.16 for Band 6 and 0.29 for Band 7 and with the exception of Band 6, the amount of variance explained by each variable increases progressively with band wavelength. These results suggest that the bands which record reflectance in the longer wavelength portion of the spectrum (TM bands 4-7) may contain more information relating to percent surface cover than TM bands 1-3. These results support previous studies which investigated the information content of TM data and found that bands 7, 5, and 4 were the optimal combination for general surface cover mapping (Horler, 1986).

Relationships between topographic variables and sensor data (4-3b) are moderate for all TM Bands with r^2 always greater than or equal to 0.29 (Band 6) and the most significant correlation with Band 5 (r^2=0.46). Band 2 and Band 3 each had correlation coefficients of $r^2 = 0.45$. These results are consistent with previous studies which have identified significant topographic effects in remotely sensed data of mountainous terrain and they support the bivariate results discussed earlier (Hall-Konyves, 1987; Dottavio, 1981).

The relationships between all ground variables and the sensor data Table 4-3(c) are always greater than or equal to $r^2 = 0.32$ (Band 6) with a

Table 4-3: Summary of Regression Results (α=0.01, DF=773)

(a) Surface Cover Variables vs Sensor Band

Band 1 : $r^2 = 0.19$, $F = 44.49$			
parameter	estimate	T^*	std. error
intercept	59.29	37.54	1.58
coniferous	0.07	3.38	0.02
deciduous	0.13	7.13	0.02
non-veg	0.19	9.94	0.02
herb	0.15	7.35	0.02
moss	0.00	.	.

Band 2 : $r^2 = 0.23$, $F = 58.27$			
intercept	20.85	19.61	1.05
coniferous	0.05	3.57	0.01
deciduous	0.10	8.16	0.01
non-veg	0.13	10.45	0.01
herb	0.12	9.01	0.01
moss	0.00	.	.

Band 3 : $r^2 = 0.24$, $F = 61.73$			
intercept	17.98	11.85	1.52
coniferous	0.05	*	0.02
deciduous	0.12	6.66	0.02
non-veg	0.19	10.36	0.02
herb	0.16	8.31	0.02
moss	0.00	.	.

Band 4 : $r^2 = 0.25$, $F = 63.73$			
parameter	estimate	T^*	std. error
intercept	33.00	9.11	3.62
coniferous	0.16	3.43	0.05
deciduous	0.41	9.82	0.04
non-veg	0.24	5.64	0.04
herb	0.46	10.03	0.05
moss	0.00	.	.

Band 5 : $r^2 = 0.27$, $F = 72.35$			
intercept	38.58	8.56	4.51
coniferous	0.18	3.11	0.06
deciduous	0.42	7.97	0.05
non-veg	0.53	9.92	0.05
herb	0.59	10.32	0.06
moss	0.00	.	.

Band 6 : $r^2 = 0.16$, $F = 35.62$			
intercept	105.38	72.49	1.45
coniferous	0.21	11.35	0.02
deciduous	0.19	11.36	0.02
non-veg	0.17	9.81	0.02
herb	0.19	10.46	0.02
moss	0.00	.	.

Band 7 : $r^2 = 0.29$, $F = 79.36$			
parameter	estimate	T^*	std. error
intercept	13.15	6.79	1.94
coniferous	0.08	3.26	0.02
deciduous	0.15	6.89	0.02
non-veg	0.29	13.49	0.02
herb	0.22	9.13	0.02
moss	0.00	.	.

* - denotes correlation not significant at 0.01 level of confidence

Table 4-3, continued

(b) Topographic Variables vs Sensor Band

Band 1 : $r^2 = 0.35$, $F = 103.33$

parameter	estimate	T*	std. error
intercept	44.75	33.72	1.33
elevation	0.01	7.24	0.00
slope	0.06	3.08	0.02
incidence	14.92	12.10	1.23
relief	0.10	7.24	0.01

Band 2 : $r^2 = 0.45$, $F = 154.46$

parameter	estimate	T*	std. error
intercept	9.24	11.00	0.84
elevation	0.01	10.56	0.00
slope	0.05	3.73	0.01
incidence	11.06	14.16	0.78
relief	0.06	6.98	0.01

Band 3 : $r^2 = 0.45$, $F = 155.07$

parameter	estimate	T*	std. error
intercept	0.44	*	1.22
elevation	0.01	12.51	0.00
slope	0.04	2.06	0.02
incidence	13.96	12.33	1.13
relief	0.10	7.21	0.01

Band 4 : $r^2 = 0.35$, $F = 103.45$

parameter	estimate	T*	std. error
intercept	3.55	*	3.17
elevation	0.02	11.50	0.00
slope	0.14	2.90	0.05
incidence	32.91	11.19	2.94
relief	-0.03	*	0.03

Band 5 : $r^2 = 0.46$, $F = 163.06$

parameter	estimate	T*	std. error
intercept	-8.23	*	3.65
elevation	0.04	16.7	0.0
slope	-0.03	*	0.06
incidence	36.95	10.89	3.39
relief	0.13	3.18	0.04

Band 6 : $r^2 = 0.29$, $F = 77.25$

parameter	estimate	T*	std. error
intercept	105.27	83.87	1.26
elevation	0.00	-2.62	0.00
slope	0.03	*	0.02
incidence	17.71	15.18	1.17
relief	0.07	5.46	0.01

Band 7 : $r^2 = 0.43$, $F = 146.15$

parameter	estimate	T*	std. error
intercept	-8.70	-5.34	1.63
elevation	0.01	11.95	0.00
slope	0.07	2.84	0.03
incidence	10.04	12.59	1.51
relief	0.10	5.89	0.02

* - denotes correlation not significant at 0.01 level of confidence

Table 4-3, continued

(c) All Ground Variables vs Sensor Band

Band 4 : $r^2 = 0.42$, $F = 69.78$			
parameter	estimate	T^*	std. error
intercept	-9.20	*	4.37
elevation	0.02	8.27	0.00
slope	0.22	4.46	0.05
incidence	26.00	8.40	3.10
relief	-0.01	*	0.03
coniferous	0.19	3.79	0.05
deciduous	0.27	6.63	0.04
non-veg	0.03	*	0.04
herb	0.21	4.56	0.05
moss	0.00	.	.
Band 5 : r^2 = 0.46==8, $F = 87.90$			
intercept	-14.71	-2.80	5.25
elevation	0.03	10.88	0.00
slope	-0.09	*	0.06
incidence	31.62	8.51	3.72
relief	0.11	2.92	0.04
coniferous	0.15	*	0.06
deciduous	0.19	3.89	0.05
non-veg	0.25	5.11	0.05
herb	0.18	3.37	0.05
moss	0.00	.	.
Band 6 : $r^2 = 0.32$, $F = 45.21$			
intercept	98.80	55.12	1.79
elevation	0.00	*	0.00
slope	0.02	*	0.02
incidence	14.38	11.32	1.27
relief	0.07	5.36	0.01
coniferous	0.11	5.42	0.02
deciduous	0.10	5.98	0.02
non-veg	0.09	5.41	0.02
herb	0.09	4.93	0.02
moss	0.00	.	.

Band 1 : $r^2 = 0.39$, $F = 61.13$			
parameter	estimate	T^*	std. error
intercept	45.21	24.03	1.88
elevation	0.00	4.31	0.00
slope	0.02	*	0.02
incidence	14.02	10.52	1.33
relief	0.10	7.14	0.01
coniferous	0.01	*	0.02
deciduous	0.04	*	0.02
non-veg	0.08	4.81	0.02
herb	0.02	*	0.02
moss	0.00	.	.
Band 2 : $r^2 = 0.47$, $F = 85.33$			
intercept	8.79	7.32	1.20
elevation	0.00	6.64	0.00
slope	0.03	*	0.01
incidence	10.19	11.98	0.85
relief	0.06	6.88	0.01
coniferous	0.01	*	0.01
deciduous	0.04	3.12	0.01
non-veg	0.05	4.71	0.01
herb	0.02	*	0.01
moss	0.00	.	.
Band 3 : $r^2 = 0.48$, $F = 88.13$			
intercept	0.68	*	1.73
elevation	0.01	8.18	0.00
slope	0.00	*	0.02
incidence	13.22	10.79	1.22
relief	0.09	6.97	0.01
coniferous	0.01	*	0.02
deciduous	0.03	*	0.02
non-veg	0.08	4.92	0.02
herb	0.02	*	0.02
moss	0.00	.	.

Band 7 : $r^2 = 0.50$, $F = 95.44$			
parameter	estimate	T^*	std. error
intercept	-7.48	-3.34	2.24
elevation	0.01	7.17	0.00
slope	0.00	*	0.03
incidence	18.13	11.44	1.58
relief	0.099	5.44	0.02
coniferous	0.01	*	0.03
deciduous	0.03	*	0.02
non-veg	0.15	7.26	0.02
herb	0.04	1*	0.02
moss	0.00	.	.

* - denotes correlation not significant at 0.01 level of confidence

maximum $r^2 = 0.50$ for Band 7, $r^2 = 0.48$ for Band 3 and $r^2 = 0.47$ for Band 2. These results indicate that the r^2 value improves little when surface cover is added beyond topographic variables alone.

Discussion:

The combined influence of topographic variables measured at a site and each TM band was significantly stronger than the corresponding influence of the surface cover percentages for all TM bands. Moderate relationships between the ground variables and sensor data, at best, suggests that the variance in TM bands is not fully accounted for by all the ground variables measured at each site. The remaining variance is possibly a function of surface type. Walsh (1987) entered nominal variables such as soil type into a regression model using a technique which employs dummy variables and obtained significantly higher overall correlations than models which did not employ these variables. More importantly, it provides support for the notion that there is variance in surface cover and topography that is not contained in any of the TM bands alone.

Summary:

The most important results of this regression technique, in addition to supporting the bivariate analysis, are: (i) the same topographic effect on the spectral data (roughly 50% explanation) was identified as in the literature for mountainous areas; (ii) no improvement (or little) was found when surface cover was added; and (iii) there was less correlation when surface cover was analysed on the spectral data. This means that topography influences spectral response to a very large degree; therefore, we must either reduce the topographic effect, if only surface cover is needed or integrate ancillary topographic information in landscape or biophysical classification where topography is critical. Also, topography explains much of the variance in surface cover so where spectral signature is not unique, topography can help or assist in separating that parcel of land from others.

4.4. Canonical Analysis

A canonical correlation analysis is a general form of regression in which the structural relationships between two data sets can be studied (Clarke, 1975). This is different from considering individual bivariate relationships or considering the combined effect of one data set on a single variable. In the context of this study, structural canonical analysis has been done to identify common patterns in the various data sets which can be interpreted in terms of the common landscape components linking the two data sets.

Vectors are extracted from each set of variables to represent the maximum variance within the sets and at the same time to maximize the correlation or shared variance between the two sets. The first set of vectors extracted is referred to as the first canonical vector pair. The second canonical vector pair is extracted to represent maximum variance that remains after the first vectors have been removed. In essence, canonical correlation is a summary analysis used to describe shared variance between two data sets.

In this research canonical analysis is used to summarize the variance shared between the following sets of variables: (i) the set of surface cover variables and the set of seven Landsat TM bands, (ii) the set of topographic variables and the set of TM bands, and (iii) the combined set of surface cover and topographic variables and the set of TM bands. The common variance extracted in these three analysis can be interpreted as representing the amount of information contained in the TM data set that describes the percent surface cover component of the landscape, the topographic character of the landscape, and the landscape system defined in terms of surface cover and morphometry, respectively.

Results:

The results of the canonical correlation analysis are presented in Table 4-4(a-c). Correlation measures/are expressed as canonical correlation coefficients (R_c). The square of the canonical correlation coefficient, the canonical coefficient of determination, describes the amount of variance shared by the two data sets. Overall correlations are significant at the .01 level of confidence based on F-tests of significance.

Table 4-4(a) summarizes the results of the canonical correlation analysis applied to the set surface cover variables and the set of TM sensor variables. Two significant vector pairs were extracted. The first pair had a canonical correlation coefficient of $R_c = .68$. Of these, the vector extracted from the sensor data set explained 44 percent of the variance in Band 7, 36 percent of the variance in Band 5, 36 percent in Band 3 and less than 30 percent in the remaining bands. Only 2 percent of the variance in the thermal band (Band 6) was accounted for. At the same time, the sensor vector accounted for 36 percent of the variance in percent coniferous cover, 18 percent in non-vegetated cover, and 13 percent in the herbaceous data. The variable extracted from the percent surface cover data set explains 74 percent of the variance in percent coniferous cover, 37 percent in non-vegetated cover, and 27 percent in herbaceous cover while simultaneously accounting for 20 percent of the variance in all of the TM bands, excluding the thermal channel (Band 6).

The second orthogonal vector pair (R_c=.51) extracted is composed primarily of variance remaining from Band 4 on the sensor side. The sensor vector accounts for 67 percent of this remaining variance (after the first vector pair has been extracted) and at most 20 percent of the remaining variance in the other bands.

Table 4-4(b) contains results of the canonical correlation analysis between the set of topographic variables and the set of TM sensor variables. The first pair of canonical vectors extracted have a correlation coefficient $R_c = .80$. The vector extracted from the sensor data explained 52% of the variance in Band 5, 50% of the variance in Band 3, 49% of the variance in Band 7, and less than 45% in the remaining bands with only 1% of the variance in the thermal band accounted for. The same vector explained 64% of the variance in elevation, 12%

Table 4-4: Results of Canonical Correlation
(a) TM Sensor/Surface Cover

First Vector Pair $R_c = .68$, $F = 37.52$

Sensor	r	r_z		Surf Cov	r	r_z
Band 1	.47	.33		coniferous	-.86	-.60
Band 2	.54	.37		deciduous	-.16	-.11
Band 3	.60	.41		non-veg	.61	.42
Band 4	.34	.24		herb	.52	.36
Band 5	.60	.41		moss	-.04	-.03
Band 6	-.13	-.09				
Band 7	.66	.45				

Second (Orthogonal) Vector Pair $R_c = .51$, $F = 24.28$

Sensor	r	r_z		Surf Cov	r	r_z
Band 1	.21	.11		coniferous	-.40	-.21
Band 2	.35	.18		deciduous	.63	-.33
Band 3	.25	.13		non-veg	-.57	-.30
Band 4	.82	.43		herb	.48	.25
Band 5	.45	.23		moss	-.38	-.20
Band 6	.32	.17				
Band 7	.15	.08				

R_c = Canonical Correlation Coefficient

r = Correlation between the variable and the canonical vector composed of a linear function of variables from the same same data set.

r_z = Correlation between the variable and the canonical vector composed of a linear function of variables from the other data set.

Table 4-4, continued
(b) TM Sensor/Topography

First Vector Pair $R_c = .80$, $F = 59.88$

Sensor	r	r_z		Topo	r	r_z
Band 1	.56	.45		elevation	.99	.80
Band 2	.67	.54		slope	.44	.35
Band 3	.71	.57		incidence	.21	.17
Band 4	.65	.53		relief	.26	.21
Band 5	.72	.63				
Band 6	.10	.08				
Band 7	.70	.57				

Second (Orthogonal) Vector Pair $R_c = .60$, $F = 31.56$

Sensor	r	r_z		Topo	r	r_z
Band 1	.61	.37		elevation	-.11	-.07
Band 2	.62	.38		slope	-.03	-.02
Band 3	.54	.33		incidence	.83	.51
Band 4	.40	.24		relief	.35	.21
Band 5	.39	.24				
Band 6	.82	.50				
Band 7	.54	.33				

R_c = Canonical Correlation Coefficient

r = Correlation between the variable and the canonical vector composed of a linear function of variables from the same data set.

r_z = Correlation between the variable and the canonical vector composed of a linear function of variables from the other data set.

Table 4-4, continued

(c) TM Sensor/Topography + Surface Cover

First Vector Pair $R_c = .81$, $F = 36.60$

Sensor	r	r_z		Top+Sur	r	r_z
Band 1	.58	.47		elevation	.98	.80
Band 2	.69	.56		slope	.44	.36
Band 3	.73	.59		incidence	.20	.17
Band 4	.62	.50		relief	.27	.22
Band 5	.78	.63		coniferous	-.74	-.61
Band 6	.10	.08		deciduous	.01	.01
Band 7	.73	.59		non-veg	.36	.29
				herb	.52	.42
				moss	-.16	-.13

Second (Orthogonal) Vector Pair $R_c = .63$, $F = 22.58$

Sensor	r	r_z		Top+Sur	r	r_z
Band 1	.55	.35		elevation	-.10	-.06
Band 2	.58	.37		slope	-.08	-.05
Band 3	.48	.31		incidence	.81	.52
Band 4	.52	.32		relief	.28	.18
Band 5	.40	.26		coniferous	.16	.10
Band 6	.82	.52		deciduous	.37	.23
Band 7	.46	.29		non-veg	-.12	-.08
				herb	-.09	-.06
				moss	-.60	-.38

R_c = Canonical Correlation Coefficient

r = Correlation between the variable and the canonical vector composed of a linear function of variables from the same data set.

r_z = Correlation between the variable and the canonical vector composed of a linear function of variables from the other data set.

of the variance in slope, 7% of the variance in relief, and 4% of the variance in incidence. On the other side of the correlation, the vector extracted from the topographic data set explained primarily variance in elevation (98%). It also accounted for 19% of the variance in slope and 4% and 7% of the variance in incidence and relief respectively. The same vector accounted for 40% of the variance in Band 5 and at least 20% of the variance in all other bands except the thermal band for which less than 1% of the variance was explained. The orthogonal vector pair extracted from the remaining variance had a correlation coefficient $R_c = .60$. The vector extracted from the sensor data contained variance primarily from variance in the thermal channel. Of the remaining variance in Band 6 after the first vector had been extracted, 67% was explained by the second vector. At least 15% of the remaining variance in the spectral bands was also explained by this vector, 26% of the remaining variance in incidence, and 12% of the remaining variance in relief.

Table 4-4c contains results of the canonical correlation analysis between the set of topographic and surface cover variables and the set of TM sensor variables. When the topographic and surface cover variables are considered as a group, the overall correlation coefficient for the first vector pair extracted is $R_c = .81$ From the sensor variables, again the first vector extracted explains variance primarily in the spectral bands while for the second vector extracted, where the overall correlation coefficient is $R_c = .63$, the thermal channel is best represented. From the ground data set composed of both the topographic and surface cover variables, elevation contributes the greatest variance (64%) to the first vector. The second vector also represents the greatest amount of variance in a topographic variable, namely incidence, for which 64% of the remaining variance is accounted for.

Discussion:

Results of the canonical analysis between the surface cover variables and the sensor data indicate that only 46% of the variance in the two data sets is shared. Therefore, terrain components other than percent surface cover of the various surface cover types influences TM response in mountainous terrain. The previous bivariate correlation and multiple regression analyses indicate that topographic variables make up an important component.

The canonical analysis between the the set of topographic variables and sensor data reveal a common variance of 64%. This supports and confirms the earlier statistics and shows the potential importance of the ancillary topographic data set for applications such as terrain classification. If the two data sets shared exactly the same variance, no new information could be provided by inclusion of the topographic data set and using either data set for classification should produce the same results. Since, however, the topographic data set contains large amounts of variance not represented in the sensor data, this data set may be used to complement the sensor data. The conclusion that arises from these results supports the main hypothesis of this thesis, that the incorporation of ancillary topographic data in the analysis of TM data should improve classification results of landscape units defined in terms of landform and surface cover. This hypothesis will be tested in the next two chapters where combined data sets will be used in integrated classification procedures.

Results of the canonical analysis between the set of sensor variables the the combined set of topographic and surface cover variables shows that the relationship between the ground variables and sensor data does not improve when surface cover variables are included in the model over when topographic variables are employed alone. This indicates that the percent surface cover data set contains little variance that is not already contained in the topographic and sensor data sets combined. Band 5 and Band 6 are the dominate variables for the first and second vectors extracted from the sensor data respectively. Elevation and incidence are dominant from the ground variable side regardless of whether or not the surface cover data are included. Since little new information is provided by the addition of the surface cover variables, the data set composed of the sensor data and the topographic data set should contain sufficient information to discriminate among landscape classes described in terms of geomorphometry and percent cover of the various vegetation types.

Summary:

The preceding canonical analysis reveals that the sensor data set contains variance that is also contained in the surface cover data set. Since, however, this common variance is relatively small, this suggests that the sensor data do not contain sufficient information to discriminate among terrain classes described in terms of the percent cover of the various surface cover types. Secondly, the sensor data set contains variance that is also contained in the topographic data set. Although the amount of variance shared is greater than the amount shared between the sensor and surface cover sets, the common variance is not 100%. Consequently, the sensor data does not contain sufficient information to discriminate among terrain classes defined in terms of elevation, slope, incidence and relief. On the other hand, this also means that the topographic data set contains variance not provided by the sensor data. Finally, when the common variance between the sensor data set and tbe combined surface cover and topographic data sets is examined, the addition of the surface cover data does not increase the shared variance by a significant amount over that which was obtained when the topographic data were used in the canonical model with the sensor data alone.

One interpretation based on these results is that the sensor data alone do not contain sufficient information to discriminate between terrain classes defined in terms of (i) percent of the various surface cover types, or (ii) topographic characteristics. It is therefore probable, that the sensor data would not be suitable for discriminating among classes defined in terms of a combination of these in the integrated approach. This hypothesis which applies to mountainous terrain, is supported by the fact that accuracy levels of terrain classifications reported in the literature by others using Landsat data alone in high relief terrain were poor. Further, the results discussed here show that additional variance is contained in topographic and surface cover data sets which may be used to add discriminatory power if integrated with the sensor data for terrain classification. A minimal improvement in overall correlation when surface cover variables were added to the topographic data set suggests that little or no additional information is contained in the surface cover data set that is not already contained in the combined sensor and topographic data set. This leads to the conclusion that an integrated sensor and topographic data set should be employed in terrain classification in regions such as this having high topographic variability. Further, an ancillary topographic data set is more readily available since it can be derived from a digital elevation model relatively easily. An ancillary data set containing percent cover values is impossible to acquire for the entire study area since it would have to be measured in the field for every pixel.

4.5. Chapter Summary

There are significant relationships between the various ground cover values measured in the field or derived from field measurements (elevation, slope incidence, and relief) and data recorded by the Landsat Thematic Mapper. These relationships expressed as correlation coefficients suggest that the land cover and topographic characteristics of the landscape are linked and that parameters of both these components have an effect on TM data. A multiple regression analysis showed this link, and by examining the effect of the set of topographic variables and the set of surface cover variables on each TM band, indicated that neither the variance in the surface cover nor the variance in the topography was fully explained by any TM band alone. The sensor variables were also considered as a set and the common variances between the sets of TM bands, topographic variables, and surface cover variables were examined. Canonical correlation coefficients were interpreted to mean that the variance in the sensor data set was not fully explained by the variance in either the surface cover, topographic, or combined data sets. This suggests that additional information may be contained in the ground variables which is not contained in the sensor data and may be useful for improving analysis results in high relief terrain.

Obtaining continuous data representing the surface cover throughout an area of an appreciable size is virtually impossible, therefore, an ancillary surface cover data set cannot be obtained. However, since it is a relatively simple procedure to acquire topographic data for the complete study area from an elevation model, the inclusion of an ancillary topographic data set in terrain analysis is suggested. The canonical correlation results further support the use of topographic data as ancillary data since there was little improvement in the overall correlation coefficient when the additional surface cover data was employed in the model over when the topographic data alone was used with the TM. The incorporation of ancillary data sets in the classification of TM data is examined in the next two chapters in order to determine the level of improvement that can be achieved by integrating data sets, particularly, TM sensor data and topography.

Chapter 5
Terrain Classification I: Discriminant Analysis of Site Data

5.1. Introduction

In this chapter, linear discriminant analysis, one form of classification based on statistical analysis of site data, is used to investigate the power of variables for correctly classifying the site data. The discriminant procedure uses information for pixels about which the class is known to generate discriminant functions. These functions can then be used to classify the original training pixels (used to generate the functions) or additional pixels for which the class in which they belong is known. One output of the discriminant analysis is a contingency table which identifies the class in which the pixel belongs according to the field data and the one in which it was assigned using the discriminant functions. In this way, the efficiency of discriminant functions and the power of variables used to develop these discriminant functions for correctly classifying the site data can be assessed.

The first step was to develop a classification scheme for the Yukon study area. Based on the exhaustive correlation and regression analysis in the previous chapter which showed a direct link between surface cover and geomorphometry in the study area, an integrated classification approach was selected for this research. Classes are defined in terms of landform, vegetation, and to a lesser degree soils characteristics. This approach is similar to that employed by Parks Canada in Kluane National Park adjacent to this study area and is a recognized systematic terrain classification system (Christian, 1958; Christian and Stewart, 1968; Robinove, 1979; Bastedo and Theberge, 1983; Franklin, 1987; Satterwhite et al., 1984) that is eminently suited for applications using remotely sensed data.

Table 5-1 outlines the classification scheme which was developed based on the various terrain types identified in the field and represented by the site data. Note that the exact description of each terrain class depends on subjective decisions made by the researcher and may vary depending on the experience of the researcher and/or the ultimate use of the classification. However, similar classes would have been chosen for any biophysical landscape classification. The correlation analysis indicated that a single land cover or vegetation classification would be unsuccessful. The results point to an integrated classification because the TM data contain information about vegetation and landform. Since, however, there is a lot of topographic variance not explored or contained in the TM data, it is necessary to input topographic information derived independently. Topography helps to explain the vegetation so that in areas that are not spectrally unique, topography will help separate the differences. It is recognized that in similar environments recent research has focussed on derivation of DEMs from the TM data themselves; but this approach was not available without extensive software development for use in the study reported here.

The 774 pixels used in the preceding correlation analysis and for which ground variables had been measured in the field were assigned to one of the terrain classes. From these data, a random subset was selected for generating the discriminant functions. The remaining pixels were kept for testing the functions. This procedure was repeated twice in order to ensure that the training and test pixels adequately represented the variance in the data sets. In each case, nine sets of functions were generated based on: (i) the TM data alone, (ii) the topographic data alone, (iii) TM + elevation, (iv) TM + slope, (v) TM + incidence, (vi) TM + relief, (vii) TM + elevation, slope, and incidence, (viii) TM + all topographic variables, (ix) TM + all ground variables measured in the field.

In generating the discriminant functions, each class is characterized in terms of their mean vector and covariance matrices. In doing so, normality is assumed. According to Swain and Davis (1978) this is a reasonable assumption for remote sensing applications such as classification in that classifiers designed

Table 5-1: Landscape Classes - Southwest Yukon Study Site

Class	Name	Description
1	Forest/Plain	- > 40 percent tree cover; black spruce dominant; occurs on alluvial plains at elevations < 1000 m and slopes < 30 degrees; medium drainage
2	Organic Terrain	- peat bog; < 10% tree coverage; stressed black spruce due to poor drainage; dominance of grass and sedge; occurs on alluvial plains; slopes < 5 degrees
3	Upland Shrub	- willow and dwarf birch shrubs dominate the vegetation; 0.1-3m height; occurs on mountain slopes 5-40 degrees; drainage is good
4	Alpine Meadow	- grasses and sedge dominate; mountain slopes at elevations 1400-1800 m; southfacing slopes; well drained
5	Alpine Tundra	- bryoid mat consisting of mosses and lichen; occurs on northfacing mountain slopes at elevations > 1100 m; very homogeneous; good drainage
6	Mountain Ridge	- top of mountain ridges; elevations > 1600 m; mosses; grasses and sedge < 0.1m height
7	Exposed Hillslope	- eroded hillslopes; exposed soil; sandy and brown in colour; patches of sedge and low shrubs; slopes > 30 degrees; elevations > 1400m
8	Valley Forest	- spruce on lower mountain slopes at elevations > 1000 m; 10-30 degrees; > 30% tree coverage; good drainage
9	Deciduous Shrub	- dominance of deciduous vegetation; balsam poplar and willow; tree height > 3m; occurs in stream beds; slopes are variable
10	Immature Spruce Forest	- < 20% black spruce coverage; tree height < 10m; stunted due to poor drainage; slopes < 15 degrees

for remote sensing applications are found to be robust. In other words classification accuracy is not very sensitive even to moderately severe violations of this assumption. (see also Tom and Miller, 1984).

5.2. Classification Accuracy

Tables 5-2 and 5-3 provide summaries of the classification accuracy of discriminant functions generated using the two test groups. Note that classes 2 and 7 are excluded from the mean accuracies for the overall classifications. These classes did not contain a sufficient number of known pixels from the field data to accurately represent these classes (see Curran and Williamson , 1985). For the same reason, all pixels belonging to these classes were used in the generation of the functions and no pixels were assigned to the test group. It was decided to generate more accurate functions rather than to sacrifice the ability of the functions to represent the classes so that the functions could be tested. The weighted mean given is a measure of the overall accuracy with a weight applied based on the number of pixels contained in each class. It is worthwhile mentioning that these representations of accuracy rely on the assumption that known pixels were accurately identified in the field. Consequently the tables represent the level of agreement between the field classification of pixels and the assignment of a pixel to a class by the discriminant functions.

Table 5-2a shows the percent classification accuracy based on functions generated using the TM data alone, topographic data alone, and various combinations of data sets for the first group of training pixels. Confidence intervals for selected functions are presented in Appendix B. Confidence intervals indicate the degree of confidence which can be placed on each classification based on the number of sample used in the analysis. Note that the confidence intervals for the test data set are greater, consequently a greater range of accuracies are possible for the classifications based on the test data.

When the TM data are used alone, overall accuracy is only 66.5%. Classes 3 and 9 are poorly defined classes with accuracies less than 50%. Classes

Table 5-2: Summary of Classification Accuracy - Percent* Classified Accurately in Class

(a) Test Group 1 ; 672 training pixels

class	TM alone	Topography alone	TM+ elevation	TM+ slope	TM+ incidence	TM+ relief	TM+Topo -relief	TM+ Topography	TM+ ground
1	79.8	30.3	95.0	83.8	78.8	82.8	87.9	89.9	98.0
2	77.8	100.0	88.9	100.0	100.0	100.0	100.0	100.0	100.0
3	42.0	76.0	54.7	45.3	56.0	44.0	70.7	96.0	100.0
4	74.3	100.0	78.6	94.3	88.6	91.4	98.6	100.0	100.0
5	59.3	83.3	57.4	70.4	85.2	72.2	94.4	100.0	100.0
6	68.4	100.0	89.5	89.5	92.1	89.5	100.0	100.0	100.0
7	70.4	100.0	100.0	100.0	100.0	77.8	100.0	100.0	100.0
8	86.7	100.0	93.3	93.3	93.3	93.3	100.0	100.0	100.0
9	42.6	85.2	55.6	57.4	40.7	53.7	90.7	96.3	100.0
10	78.8	93.9	78.8	87.9	84.1	83.3	94.7	97.7	100.0
mean**	66.5	83.6	75.4	77.7	77.4	76.3	92.1	97.5	99.8
w.mean	65.2	80.6	74.6	75.7	76.0	73.1	89.3	96.9	99.7

(b) Test Group 1 ; 102 test pixels

class	TM alone	Topography alone	TM+ elevation	TM+ slope	TM+ incidence	TM+ relief	TM+Topo -relief	TM+ Topography	TM+ ground
1	83.3	33.3	88.9	94.4	77.8	94.4	88.9	94.4	100.0
3	14.3	66.7	42.9	42.9	33.3	19.1	57.1	100.0	100.0
4	72.7	100.0	90.9	90.9	81.8	90.9	100.0	100.0	100.0
5	55.6	66.7	66.7	55.6	66.7	77.8	88.9	100.0	100.0
6	57.1	100.0	100.0	71.4	85.7	85.7	100.0	100.0	100.0
8	83.3	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
9	55.6	88.9	55.6	55.6	33.3	55.6	100.0	100.0	100.0
10	90.5	95.2	85.7	90.5	85.7	90.5	95.2	95.2	100.0
mean	64.1	81.4	78.8	75.2	70.5	76.8	91.3	98.7	100.0
w.mean	62.7	76.5	75.5	74.5	67.6	72.5	87.3	98.0	100.0

* - denotes correlation not significant at 0.01 level of confidence

** - excludes classes 2 and 7

Table 5-3: Summary of Classification Accuracy - Percent* Classified Accurately in Class

(a) Test Group 2 ; 635 training pixels

	TM	Topography	TM+	TM+	TM+	TM+	TM+Topo	TM+	TM+
class	alone	alone	elevation	slope	incidence	relief	-relief	Topography	ground
1	80.2	31.3	94.8	85.4	80.2	84.4	91.7	92.7	100.0
2	77.8	100.0	88.9	100.0	100.0	100.0	100.0	100.0	100.0
3	42.9	75.7	55.7	45.0	50.0	45.0	69.3	97.1	100.0
4	72.7	100.0	78.8	93.9	87.9	89.4	100.0	100.0	100.0
5	57.7	94.2	61.5	71.2	82.7	75.0	96.2	100.0	100.0
6	64.9	100.0	89.2	86.5	89.2	83.8	100.0	100.0	100.0
7	70.4	100.0	100.0	100.0	100.0	77.8	100.0	100.0	100.0
8	89.3	100.0	92.9	98.4	92.9	89.3	100.0	100.0	100.0
9	44.9	85.7	61.2	59.2	42.9	61.2	87.8	95.9	100.0
10	81.2	94.3	80.3	87.7	86.1	86.1	95.1	97.6	100.0
mean**	66.7	85.2	76.8	78.2	76.5	76.8	92.5	97.9	100.0
w.mean	65.8	81.6	76.1	76.2	75.3	74.3	89.8	97.5	100.0

(b) Test Group 2 ; 139 test pixels

	TM	Topography	TM+	TM+	TM+	TM+	TM+Topo	TM+	TM+
class	alone	alone	elevation	slope	incidence	relief	-relief	Topography	ground
1	90.5	28.6	90.5	85.7	85.7	90.5	85.7	95.2	95.2
3	41.9	71.0	54.8	51.6	51.6	35.5	64.5	93.6	100.0
4	86.7	100.0	86.7	86.7	93.3	93.3	100.0	100.0	100.0
5	54.6	100.0	45.5	72.7	63.6	54.6	90.9	100.0	100.0
6	87.5	100.0	100.0	100.0	100.0	100.0	87.5	100.0	100.0
8	62.5	100.0	75.0	75.0	75.0	75.0	100.0	100.0	100.0
9	37.5	85.7	50.0	57.1	42.9	50.0	78.6	92.9	100.0
10	80.7	93.6	80.7	80.7	77.4	77.4	90.3	93.6	100.0
mean	67.7	84.9	72.9	76.2	73.7	72.0	87.2	96.9	99.4
w.mean	66.9	79.9	72.0	73.4	71.2	68.3	84.2	95.7	99.3

* - denotes correlation not significant at 0.01 level of confidence

** - excludes classes 2 and 7

1, 8, and 10 have accuracies greater than 78%. Note that these classes are defined largely on the basis of forest conditions. When the topographic data are used alone, overall accuracy increases to 83.6%. All classes improved except class 1 for which accuracy was reduced to 30.3%. Class 3 remains the poorest defined of all classes but the accuracy has increased from 42.0% to 76.0%.

Adding a single topographic variable (elevation) to the TM data set resulted in significant improvements compared to when the TM data were used alone. An overall accuracy of 75.4% was obtained. The addition of elevation improved classification results for each class except class 10 which was originally defined well and remained the same. However, results including overall accuracy were generally less than when topographic data was employed alone except for a major increase from 30.3% to 95.0% for class 1.

The addition of either slope, incidence, or relief to the TM data showed similar improvements in overall accuracy compared to the use of TM data alone. Overall accuracies were 77.7%, 77.4%, and 76.3% respectively. There were improvements in all classes with the exception of an insignificant reductions in classes 1 and 9 when incidence was used. As in the case of adding elevation, accuracies were generally less than when the topographic data were used alone except for marked increases in class 1.

In the next step, elevation, slope and incidence were added to the TM data to discriminate classes. Overall accuracy increased to 92.1%. This was above that obtained with either the TM data or the topographic data used alone. This was also an increase over any single topographic variable added to the TM data. Individual class accuracies also improved or stayed the same in all classes except for class 1 where the use of TM and elevation alone produce the best results and class 3 in which the topographic data alone best discriminated this class. The further addition of relief so that all topographic variables and all TM data was employed further improved overall accuracy to 97.5% and improved all classes further except in the case of the first exception identified above, class 1, where results were the next best to adding elevation to the TM data alone.

Generally, the same trends can be identified in Table 5-2(b) which shows similar results for the first group of test pixels. There are improvements in overall classification accuracy when either topographic variable is added to the TM data set and an even greater accuracy when the topographic data are used alone. There is a further improvement when three of the topographic variables, elevation slope, and incidence are added to the TM data set to discriminate the classes and still a greater improvement when all topographic variables are added. Trends of individual classes are also similar with classes 3 and 9 the most poorly defined using the TM data alone and class 1 with poor definition based on the topographic data alone. All classes show the best results when TM and the complete topographic data set is used.

Table 5-3 shows the results of the same procedure applied to a second group of pixels. Again, overall classification accuracies reflect similar trends when compared to the results of the first group for both the training and test pixels. Since the impact of using certain variable to discriminate among the classes is consistent regardless of which training or test group is employed and since the trained and test groups were selected randomly, this suggests that pixels selected and used to train and test each class, represent the data set well. Consequently, in spite of the relatively small number of known pixels, these results are considered to be consistent and reliable and can be used to reflect the impact of employing various TM and topographic combinations in discrimination of landscape classes in the study area.

While the class accuracies reveal the impact of incorporating topographic variables in the discrimination process, they do not suggest where the problems of confusion between classes occur. Table 5-4(a-i) contains contingency matrices constructed after using each of the nine discriminant function to separately classify the 672 training pixels and the 102 test pixels in test group 1. These tables show how pixels from each class were classified according to the discriminant functions. The vertical diagonal represents the number of pixels correctly grouped in to each class and corresponds to the percentages presented in

Table 5-2. Omission errors represent pixels identified as belonging to one class according to the field data but assigned differently by the discriminant functions. Commission errors are pixels identified as belonging to different classes from the field data but assigned to the same class by the discriminant functions. The classifications based on both the training and test data sets are provided. Similar tables were produced for the second test group but are not included because results were similar as indicated in the overall summaries provided in Table 5-2 and Table 5-3.

When the TM variables are used alone to discriminate the classes, there are significant errors of omission for all classes. The largest omission errors occurred for class 3 (Upland Shrub) where 87 of the 150 pixels assigned to that class are classified incorrectly. Most of these pixels were incorrectly assigned to class 8 (Valley Forest) or class 10 (Immature Spruce Forest) but there was confusion with all classes except class 2 (Organic Terrain). In the test data set, major confusion occurred with class 9 (Deciduous Shrub). This suggests that this class is poorly defined based on TM data and alone and more discriminatory information is required if the accuracy of this class is to be improved. Omission errors are also large for class 9 where 11 of the 54 pixels defined as belonging to class 9 were assigned to class 3. This further shows confusion between classes 3 and 9 and indicates that these classes are spectral very similar. If we consider the surface cover description of these two classes we note that both contain deciduous type vegetation. It is the topographic context that makes them distinct. The incorporation of topographic variables in the discrimination process should thus improve results. Although class 8 has relatively few errors of omission, 46 of the 72 pixels assigned to class eight are incorrectly done so in the training data set and 8 of the 13 in the test data set. . The main sources of confusion are classes 3 and 10. Again this is because class eight is spectrally not unique. Topographically and by definition it only occurs in valleys. Incorporation of topographic data should therefore, reduce commission errors and thereby improve class discrimination.

Table 5-4: Contingency Tables for Discriminant Classification

(a) TM Data Alone

Training data (N=672):

	number pixels grouped into class											
class	1	2	3	4	5	6	7	8	9	10	tot	omis
1	79	3	2	1	1	0	2	4	1	6	99	20
2	1	14	0	0	0	0	0	1	0	2	18	4
3	3	0	63	17	10	2	6	21	7	21	150	87
4	0	1	2	52	7	3	2	3	0	0	70	18
5	1	3	1	0	32	3	2	3	1	8	54	22
6	1	2	0	0	3	26	5	0	1	0	38	12
7	0	1	0	0	2	2	19	1	0	2	27	8
8	1	0	0	0	0	0	0	26	0	3	30	4
9	6	2	11	0	4	0	1	4	23	3	54	31
10	7	4	2	0	4	0	2	9	0	104	132	28
total	99	30	81	70	63	36	39	72	33	149	672	234
comm	20	16	18	18	31	10	20	46	10	45	234	

Test Data (N=102):

	number pixels grouped into class											
class	1	2	3	4	5	6	7	8	9	10	tot	omis
1	15	1	0	0	0	0	1	0	0	1	18	3
3	1	0	3	4	1	0	1	4	3	4	21	18
4	0	0	0	8	0	0	1	1	1	0	11	3
5	0	0	1	0	5	0	2	0	0	1	9	4
6	0	0	0	0	0	4	2	1	0	0	7	3
8	0	0	0	0	0	0	0	5	0	1	6	1
9	1	1	1	0	0	0	0	1	5	0	9	4
10	1	0	0	0	0	0	0	1	0	19	21	2
total	18	2	5	12	6	4	7	13	9	26	102	38
comm	3	2	2	4	1	0	7	8	4	7	38	

(b) Topographic Data Alone

Training Data (N=672):

	number pixels grouped into class											
class	1	2	3	4	5	6	7	8	9	10	tot	omis
1	30	0	0	0	0	0	0	0	0	69	99	69
2	0	18	0	0	0	0	0	0	0	0	18	0
3	0	0	114	8	10	2	0	0	16	0	150	36
4	0	0	0	70	0	0	0	0	0	0	70	70
5	0	0	0	0	45	0	0	0	2	7	54	9
6	0	0	0	0	0	38	0	0	0	0	38	0
7	0	0	0	0	0	0	27	0	0	0	27	0
8	0	0	0	0	0	0	0	30	0	0	30	0
9	0	0	0	0	0	0	0	0	46	8	54	8
10	0	0	8	0	0	0	0	0	0	124	132	8
total	30	18	122	78	55	40	27	30	64	208	672	130
comm	0	0	8	8	10	2	0	0	18	84	130	

Test Data (N=102):

	number pixels grouped into class											
class	1	2	3	4	5	6	7	8	9	10	tot	omis
1	6	0	0	0	0	0	0	0	0	12	18	12
3	0	0	14	1	2	0	0	0	4	0	21	7
4	0	0	0	11	0	0	0	0	0	0	11	0
5	0	0	0	0	6	0	0	0	1	2	9	3
6	0	0	0	0	0	7	0	0	0	0	7	0
8	0	0	0	0	0	0	0	6	0	0	6	0
9	0	0	0	0	0	0	0	0	8	1	9	1
10	0	0	1	0	0	0	0	0	0	20	21	1
total	6	0	15	12	8	7	0	6	13	35	102	24
comm	0	0	1	1	2	0	0	0	5	15	24	

Table 5-4, continued

(c) TM + Elevation

Training Data (N=672):

	number pixels grouped into class											
class	1	2	3	4	5	6	7	8	9	10	tot	omis
1	94	0	0	0	0	0	0	1	0	4	99	5
2	2	16	0	0	0	0	0	0	0	0	18	2
3	0	0	82	11	15	0	3	20	12	7	150	68
4	0	0	4	55	5	6	0	0	0	0	70	15
5	0	0	1	2	31	0	8	4	0	8	54	23
6	0	0	0	0	0	34	4	0	0	0	38	4
7	0	0	0	0	0	0	27	0	0	0	27	0
8	0	0	0	0	1	0	0	28	0	1	30	2
9	0	2	10	0	5	0	0	3	30	4	54	24
10	15	0	0	0	7	0	0	6	0	104	132	28
total	111	18	97	68	64	40	42	62	42	128	672	171
comm	17	2	15	13	33	6	15	34	12	24	171	

Test Data (N=102):

	number pixels grouped into class											
class	1	2	3	4	5	6	7	8	9	10	tot	omis
1	16	1	0	0	0	0	0	0	0	1	18	2
3	0	0	9	2	3	0	1	1	3	2	21	12
4	0	0	0	10	1	0	0	0	0	0	11	1
5	0	0	0	0	6	0	1	1	1	0	9	3
6	0	0	0	0	0	7	0	0	0	0	7	0
8	0	0	0	0	0	0	0	6	0	0	6	0
9	0	1	1	0	1	0	0	1	5	0	9	4
10	2	0	0	0	1	0	0	0	0	18	21	3
total	18	2	10	12	12	7	2	9	9	21	102	25
comm	2	2	1	2	6	0	2	3	4	3	25	

(d) TM + Slope

Training Data (N=672):

	number pixels grouped into class											
class	1	2	3	4	5	6	7	8	9	10	tot	omis
1	83	0	3	1	0	0	0	1	1	10	99	16
2	0	18	0	0	0	0	0	0	0	0	18	0
3	4	0	68	25	7	1	0	21	12	12	150	82
4	0	0	0	66	4	0	0	0	0	0	70	4
5	0	0	0	0	38	4	0	0	5	7	54	16
6	1	0	0	0	3	34	0	0	0	0	38	4
7	0	0	0	0	0	0	27	0	0	0	27	0
8	0	0	0	0	1	0	0	28	0	1	30	2
9	2	0	8	0	2	0	0	9	31	2	54	23
10	5	0	3	0	0	0	0	7	1	116	132	16
total	95	18	82	92	55	39	27	68	50	148	672	163
comm	12	0	14	26	17	5	0	38	19	32	163	

Test Data (N=102):

	number pixels grouped into class											
class	1	2	3	4	5	6	7	8	9	10	tot	omis
1	17	0	0	0	0	0	0	0	0	1	18	1
3	0	0	9	2	2	0	0	2	3	3	21	12
4	0	0	0	10	0	0	0	0	1	0	11	1
5	0	0	1	0	5	1	0	1	0	1	9	4
6	0	0	0	1	0	5	0	1	0	0	7	2
8	0	0	0	0	0	0	0	6	0	0	6	0
9	1	0	2	0	0	0	0	1	5	0	9	4
10	1	0	1	0	0	0	0	0	0	19	21	2
total	19	0	13	13	7	6	0	11	9	24	102	26
comm	2	0	4	3	2	1	0	5	4	5	26	

Table 5-4, continued

(e) TM + Incidence

Training Data (N=672):

number pixels grouped into class												
class	1	2	3	4	5	6	7	8	9	10	tot	omis
1	78	0	3	2	0	0	0	3	1	12	99	21
2	0	18	0	0	0	0	0	0	0	0	18	0
3	3	0	84	9	11	8	0	20	10	5	150	66
4	0	0	3	62	3	0	0	2	0	0	70	8
5	2	0	0	0	46	1	0	0	1	4	54	8
6	1	0	0	0	1	35	0	1	0	0	38	3
7	0	0	0	0	0	0	27	0	0	0	27	0
8	2	0	0	0	0	0	0	28	0	0	30	2
9	2	0	11	0	8	0	0	6	22	5	54	32
10	5	0	1	0	0	1	0	13	1	111	132	21
total	93	18	102	73	69	45	27	73	35	137	672	161
comm	15	0	18	11	23	10	0	45	13	26	161	

Test Data (N=102):

number pixels grouped into class												
class	1	2	3	4	5	6	7	8	9	10	tot	omis
1	14	0	0	0	0	0	0	2	0	2	18	4
3	0	0	7	5	3	1	0	3	1	1	21	14
4	0	0	0	9	0	0	0	1	1	0	11	2
5	0	0	0	0	6	1	0	0	1	1	9	3
6	0	0	0	0	0	6	0	1	0	0	7	1
8	0	0	0	0	0	0	0	6	0	0	6	0
9	1	0	2	0	1	0	0	1	3	1	9	6
10	1	0	0	0	0	0	0	2	0	18	21	3
total	16	0	9	14	10	8	0	16	6	23	102	33
comm	2	0	2	5	4	2	0	10	3	5	33	

(f) TM + Relief

Training Data (N=672):

number pixels grouped into class												
class	1	2	3	4	5	6	7	8	9	10	tot	omis
1	82	1	1	2	1	0	1	5	2	4	99	17
2	0	18	0	0	0	0	0	0	0	0	18	0
3	6	3	66	10	18	0	2	23	8	14	150	84
4	0	0	0	64	4	2	0	0	0	0	70	6
5	0	0	1	0	39	1	0	1	1	11	54	15
6	1	0	0	0	1	34	1	1	0	0	38	4
7	0	0	0	0	3	0	21	1	0	2	27	6
8	2	0	0	0	0	0	0	28	0	0	30	2
9	6	0	5	0	3	0	4	5	29	2	54	25
10	5	0	3	0	3	0	3	7	1	110	132	22
total	102	22	76	76	72	37	32	71	41	143	672	181
comm	20	4	10	12	33	3	11	43	12	33	181	

Test Data (N=102):

number pixels grouped into class												
class	1	2	3	4	5	6	7	8	9	10	tot	omis
1	17	0	0	0	0	0	0	0	0	1	18	1
3	3	0	4	4	3	0	0	4	1	2	21	17
4	0	0	0	10	0	0	0	0	1	0	11	1
5	0	0	1	0	7	0	0	1	0	0	9	2
6	0	0	0	0	0	6	0	1	0	0	7	1
8	0	0	0	0	0	0	0	6	0	0	6	0
9	1	0	0	0	0	0	1	2	5	0	9	4
10	1	0	0	0	0	0	0	1	0	19	21	2
total	22	0	5	14	10	6	1	15	7	22	102	28
comm	5	0	1	4	3	0	1	9	2	3	28	

Table 5-4, continued

(g) TM + Elevation, Slope, and Incidence

Training Data (N=672):

number pixels grouped into class												
class	1	2	3	4	5	6	7	8	9	10	tot	omis
1	87	0	0	0	0	0	0	0	0	12	99	12
2	0	18	0	0	0	0	0	0	0	0	18	0
3	0	0	106	15	9	0	0	3	11	6	150	44
4	0	0	1	69	0	0	0	0	0	0	70	1
5	0	0	0	0	51	0	0	0	0	3	54	3
6	0	0	0	0	0	38	0	0	0	0	38	0
7	0	0	0	0	0	0	27	0	0	0	27	0
8	0	0	0	0	0	0	0	30	0	0	30	0
9	0	0	3	0	0	0	0	1	49	1	54	5
10	5	0	2	0	0	0	0	0	0	125	132	7
total	92	18	112	84	60	38	27	34	60	147	672	72
comm	5	0	6	15	9	0	0	4	11	22	72	

Test Data (N=102):

number pixels grouped into class												
class	1	2	3	4	5	6	7	8	9	10	tot	omis
1	16	0	0	0	0	0	0	0	0	2	18	2
3	0	0	12	3	3	0	0	0	3	0	21	9
4	0	0	0	11	0	0	0	0	0	0	11	0
5	0	0	0	0	8	0	0	0	1	0	9	1
6	0	0	0	0	0	7	0	0	0	0	7	0
8	0	0	0	0	0	0	0	6	0	0	6	0
9	0	0	0	0	0	0	0	0	9	0	9	0
10	1	0	0	0	0	0	0	0	0	20	21	1
total	17	0	12	14	11	7	[illegible]	6	13	22	102	13
comm	1	0	0	3	3	0	0	0	4	2	13	

(h) TM + All Topography

Training Data (N=672):

number pixels grouped into class												
class	1	2	3	4	5	6	7	8	9	10	tot	omis
1	89	0	0	0	0	0	0	0	0	10	99	10
2	0	18	0	0	0	0	0	0	0	0	18	0
3	0	0	144	1	1	0	0	0	1	3	150	6
4	0	0	0	70	0	0	0	0	0	0	70	0
5	0	0	0	0	54	0	0	0	0	0	54	0
6	0	0	0	0	0	38	0	0	0	0	38	0
7	0	0	0	0	0	0	27	0	0	0	27	0
8	0	0	0	0	0	0	0	30	0	0	30	0
9	0	0	1	0	0	0	0	0	52	1	54	2
10	2	0	1	0	0	0	0	0	0	129	132	3
total	91	18	146	71	55	38	27	30	53	143	672	21
comm	2	0	2	1	1	0	0	0	1	14	21	

Test Data (N=102):

number pixels grouped into class												
class	1	2	3	4	5	6	7	8	9	10	tot	omis
1	17	0	0	0	0	0	0	0	0	1	18	1
3	0	0	21	0	0	0	0	0	0	0	21	0
4	0	0	0	11	0	0	0	0	0	0	11	0
5	0	0	0	0	9	0	0	0	0	0	9	0
6	0	0	0	0	0	7	0	0	0	0	7	0
8	0	0	0	0	0	0	0	6	0	0	6	0
9	0	0	0	0	0	0	0	0	9	0	9	0
10	1	0	0	0	0	0	0	0	0	20	21	1
total	18	0	21	11	9	7	0	6	9	21	102	2
comm	1	0	0	0	0	0	0	0	0	1	2	

Table 5-4, continued

(i) TM + Ground

Training Data (N=672):

number pixels grouped into class												
class	1	2	3	4	5	6	7	8	9	10	tot	omis
1	97	0	0	0	0	0	0	0	0	2	99	2
2	0	18	0	0	0	0	0	0	0	0	18	0
3	0	0	150	0	0	0	0	0	0	0	150	0
4	0	0	0	70	0	0	0	0	0	0	70	0
5	0	0	0	0	54	0	0	0	0	0	54	0
6	0	0	0	0	0	38	0	0	0	0	38	0
7	0	0	0	0	0	0	27	0	0	0	27	0
8	0	0	0	0	0	0	0	30	0	0	30	0
9	0	0	0	0	0	0	0	0	54	0	54	0
10	0	0	0	0	0	0	0	0	0	132	132	0
total	97	18	150	70	54	38	27	30	54	134	672	2
comm	0	0	0	0	0	0	0	0	0	2	2	

Test Data (N=102):

number pixels grouped into class												
class	1	2	3	4	5	6	7	8	9	10	tot	omis
1	18	0	0	0	0	0	0	0	0	0	18	0
3	0	0	21	0	0	0	0	0	0	0	21	0
4	0	0	0	11	0	0	0	0	0	0	11	0
5	0	0	0	0	9	0	0	0	0	0	9	0
6	0	0	0	0	0	7	0	0	0	0	7	0
8	0	0	0	0	0	0	0	6	0	0	6	0
9	0	0	0	0	0	0	0	0	9	0	9	0
10	0	0	0	0	0	0	0	0	0	21	21	0
total	18	0	21	11	9	7	0	6	9	21	102	0
comm	0	0	0	0	0	0	0	0	0	0	0	

The use of topographic data alone reduces total errors of omission from 234 to 130 in the training data set and from 38 to 24 in the test data. This overall reduction is contributed to by all classes except class 1 (Forest/Plain) and class 4 (Alpine Meadow) where omission errors increase by 49% and 51% pixels respectively for the training data set and by 9% pixels for class 1 in the test data set. The increase in omission errors for class 10 can be explained by the fact that the Spruce Forest class is defined relatively well spectrally and is not discriminated well on the basis of its topography. Although omission errors are higher, however, there are no commission errors, that is, no pixels were incorrectly assigned into this class. This could reflect that the training pixels simply did not capture the full topographic variability with this class in which case the incorporation of topographic variables would enforce stricter topographic limits on the assignment of a pixels to this class than what should be imposed.

The incorporation of either elevation, slope, incidence, or relief (Table 5-4c,d,e,f) reveal decreases in the number of omission error compared to when TM data were used alone but not to the level when topographic variables are included as a group either of elevation, slope and incidence or with the inclusion of relief. The lowest overall number of omission errors (21 of the 672 training data and 2 of the 102 test pixels) occurs when all topographic variables are included and are further reduced to 2 and 0 when all ground variables including the percent surface cover data are employed. When all the topographic data are used with the TM data the omission errors that occur are primarily a result of confusion between classes 1 and 10. As mentioned previously this may be a result of the training data not encompassing a wide enough range of topographic conditions for this class. The inclusion of additional training data for this class might therefore improve this class discrimination. Since class 1 was defined better when only elevation was added to the TM data this suggests that subtle variations in incidence, relief, and slope that is not captured by the training data could largely be the cause.

5.3. Summary

In this chapter, known pixels based on field analysis of an area of mountainous terrain in the Southwest Yukon are assigned to landscape classes defined in terms of surface cover and morphometry. The classification of each pixel is dependent on discriminant functions developed based on TM data alone, topographic data alone, and combinations of TM, topographic, and surface cover information. Contingency tables developed after each discriminant function is used to classify random selections of training and test pixels are examined to determine the level of improvements that can be achieved by incorporating various topographic parameters in the classification process.

Results of this chapter have provided evidence that topographic data are necessary for terrain classification in high relief environments. The topographic data employed in the preceding analysis, however, were measured or derived from measurements taken in the field. If topographic data are to be employed in a classification of the complete study area, a continuous data set must be available. Such data are obtainable from a digital elevation model. The level of accuracy that can be expected on a more operational level using DEM derived data will be examined in the next chapter. Further, the preceding analysis uses one classification technique, discriminant analysis. Although the results of discriminant analysis are statistically precise, image analysis systems employed for producing classification results in the form of a map often use a maximum likelihood or other type of classifier. The next chapter investigates data integration from a spatial perspective using a maximum likelihood classifier to classify and map all pixels in the study area. Assessments are made based on the known field pixels.

Chapter 6

Terrain Classification II: Maximum Likelihood Classification and Mapping

6.1. Introduction

In this chapter, integrated classification is presented which incorporates spectral data acquired by Landsat and topographic data extracted from a digital elevation model. Maps of the distribution of terrain classes are included which were produced based on (i) the spectral data alone and (ii) a combined spectral and topographic data set. A qualitative assessment of the spatial effects of incorporating topographic parameters in the digital classification is then discussed. This is followed by quantitative accuracy assessments of the two maps based on site data acquired in the field and used in the discriminant analysis (discussed in the previous chapter).

A supervised classification approach was taken which involves *training* a classifier to recognize the classes of interest to the map user. In the training procedure, the image analyst identifies areas on the image that are known to represent each class from field data. The data associated with these areas are then used to develop class signatures for each class. This is one stage in which geomorphometric parameters may be incorporated into the classification. In this research, two sets of class signatures were developed from the same training areas for each class. For the first set, class signatures were developed based on spectral TM bands 1-7 alone. For the second set, elevation, slope, and incidence were included in the class signature. A maximum likelihood classification algorithm was then employed to classify each pixel in the study area into one of the terrain classes based (i) on the spectral signatures and (ii) on the integrated spectral and topographic signatures. Maps were output from the classification to show the spatial distribution of terrain classes on an Ink Jet Plotter linked to the ARIES III. Classification summary tables were also produced.

6.2. Spatial Analysis

Maps of the distribution of terrain classes are presented in Figures 6-1 and 6-2. Figure 6-1 was produced based on the spectral data alone. Figure 6-2 is the integrated classification. Summaries for each class are presented in Table 6-1a and b.

Some spatial effects of incorporating geomorphometric parameters in the classification are immediately obvious when comparing the two maps. The most evident is in the spatial distribution of class 8. On Figure 6-1, class 8 is incorrectly mapped along the alluvial plain. This class, by definition, should show up only throughout the mountain valley system. In Figure 6-2 where geomorphometric parameters have been employed to discriminate classes, the problem is eliminated and class 8 is correctly mapped only in the valleys. While areas along the plains are spectrally similar to those in the valleys due to the vegetation composition which is primarily black spruce, topographically, the two areas are distinct. This clearly illustrates that topographic information is necessary if this class is to be mapped successfully in this region.

Another variation between the two maps occurs in the southern part in an area referred to as the Burwash Flats. On Figure 6-1 a large section of this area is identified as belonging to class 6 (mountain ridge). The highest elevation in the area is 5500 feet. This is far below the lower limit of elevation for an area to be classified as mountain ridge in this area. When elevation and the other geomorphometric parameters are included in the classification, this area is correctly identified as class 4.

A third observable difference between the two maps is in the level of homogeneity of classes. In Figure 6-1, small local variations in spectral

Table 6-1: Class Summaries for Maximum Likelihood Classifications

(a) TM Data Alone

class	pixels	area	% of	class label	map colour
1	21483	193347.00	7.11	Forest Plain	Brown
2	16682	150138.00	5.52	Organic Terrain	Dark Blue
3	73971	665739.00	24.50	Upland Shrub	Light Green
4	26219	235971.00	8.68	Alpine Meadow	Yellow
5	21628	194652.00	7.16	Alpine Tundra	Pink
6	21844	196596.00	7.23	Mountain Ridge	Grey
7	24980	224820.00	8.27	Exposed Hillslope	Light Blue
8	39418	354762.00	13.05	Valley Forest	Red
9	9032	812880.00	2.99	Deciduous Shrub	Dark Green
10	33459	301131.00	11.08	Immature Spruce	Orange
uncl	13232	119088.00	4.38	Unclassified	Black

(b) Combined TM and Topographic data

class	pixels	area	% of	class label	map colour
1	24078	216702.00	7.97	Forest Plain	Brown
2	7637	68733.00	2.53	Organic Terrain	Dark Blue
3	97984	881856.00	32.45	Upland Shrub	Light Green
4	41522	373698.00	13.75	Alpine Meadow	Yellow
5	14361	129249.00	4.76	Alpine Tundra	Pink
6	13289	119601.00	4.40	Mountain Ridge	Grey
7	12611	113499.00	4.18	Exposed Hillslope	Light Blue
8	16088	144792.00	5.33	Valley Forest	Red
9	9921	89289.00	3.29	Deciduous Shrub	Dark Green
10	34282	3085538.00	11.35	Immature Spruce	Orange
uncl	30175	271575.00	9.99	Unclassified	Black

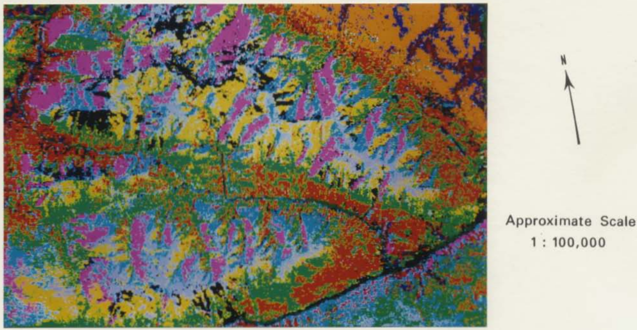


Figure 6-1: Supervised Classification of the Southwest Yukon Study Area based on TM Bands 1-7

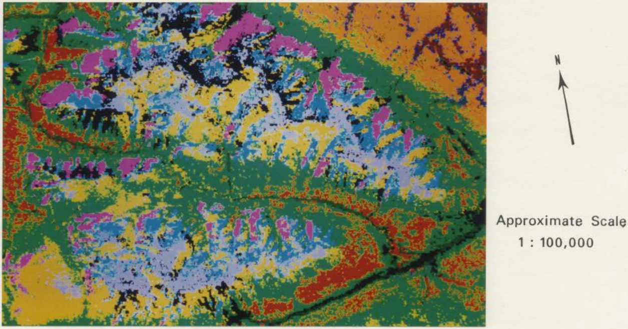


Figure 6-2: Supervised Classification of the Southwest Yukon Study Area based on TM Bands 1-7 + Elevation, Slope, and Incidence

characteristics show up as different classes. Consequently the map appears noisy. In Figure 6-2, classes appear more homogeneous. A possible explanation is that in order for a pixel to be assigned to a class in the integrated classification, the data for that pixel must satisfy both spectral and topographic conditions. The possible range of spectral values for each class will consequently be broader.

Another observation worthy of explanation which is less favourable in terms of the integrated classification is the greater percentage of unclassified pixels. This is in part because pixels must satisfy more criteria before being assigned to a class and the full range of values for each parameter must be captured during the training stage. As noted by Hutchinson (1982), this becomes more difficult as the number of parameters increases. In this study, the number of unclassified pixels could be reduced if more training data were available. Further, there may be terrain classes that were not included in the random selection of pixels visited in the field and consequently were not trained upon in the classification procedure.

6.3. Mapping Accuracy

The accuracy assessment of the digital classifications is based on the level of agreement between field classification of 774 pixels and the spectral and integrated classifications. It is worthwhile repeating from chapter 5 that the assessment relies on the assumption of accurate field classification. As a result, the contingency tables represent the level of agreement between the field data and the digital classifications rather than the level of accuracy of the maps.

The terrain class assigned to each pixel was determined according to the the following procedure. Each site was identified on the digital image classifications using the DIPIX ARIES Image Analysis System. The nine pixels which make up each site were examined to determine the dominant terrain class. The class label given to the site was simply the class in which the greatest number of pixels in the site had been assigned. All pixels within that site would then be given that class label for the purpose of assessing the agreement between field observations and the digital classification. This approach ensured a high level of locational accuracy with respect to identifying individual pixels in the field. While it was not always possible to precisely locate a single 30m x 30m pixel, locating a $8100m^2$ area which constitutes a site and includes that pixel was done with confidence.

Table 6-2 contains accuracy summaries of the maximum likelihood classifications produced based on the TM data alone and the combined TM and topographic data (elevation, slope, and incidence). Table 6-3a and 6-3b contain contingency tables for the digital classifications based on TM data alone and the integrated data sets respectively. For each table, the diagonal gives the number of pixels identified as belonging to the same class by both the digital classification and field classification. Omission errors are pixels that were identified as belonging to one class from the field data but were classified differently by the digital classification. Commission errors represent pixels classified differently from the field data but assigned to the same class in the digital classification. Confidence intervals for these tables are presented in Appendix B.

The digital classification based on the TM data alone (see Tables 6-2 and 6-3a) has an average overall classification accuracy of 55.8%. Note that classes 2 and 7 were omitted from the average as in the discriminant classifications as a result of the limited number of training data available. Average accuracies are highest for classes 1, 8, and 10. Note that these represent the forest classes which are located in areas where variations in topography are low. Although class 8 is mapped with 75% accuracy, there are however, large errors of commission. Of the 108 pixels classified as class 8 by the maximum likelihood classifier, 81 pixels actually belonged to other classes according to the field data. The spatial effects of these errors were partially identified in the previous section in that a large number of pixels on the alluvial plain were incorrectly assigned to class 8.

Table 6-2: Summary of Mapping Accuracy - Percent* Classified Accurately in Class

Class	TM Alone	TM + Topography
1	84.6	100.0
2	50.0	50.0
3	63.2	84.2
4	56.0	100.0
5	43.0	57.1
6	40.0	80.0
7	100.0	100.0
8	75.0	100.0
9	14.3	28.6
10	70.6	70.6
mean**	55.8	77.6
w.mean	61.6	79.1

* - subject to rounding error

** - excludes classes 2 and 7

Table 6-3: Contingency Tables for Maximum Likelihood Classification

(a) TM Data Alone

number pixels grouped into class													
class	1	2	3	4	5	6	7	8	9	10	uncl	tot	omis
1	99	18	0	20	0	0	0	0	0	0	0	117	18
2	0	9	0	0	9	0	0	0	0	0	0	18	9
3	0	0	108	0	0	0	9	27	27	9	0	171	63
4	0	0	9	45	9	9	9	0	0	0	0	81	36
5	0	0	0	9	27	0	9	18	0	0	0	63	36
6	0	0	0	0	9	18	18	0	0	0	0	45	27
7	0	0	0	0	0	0	27	0	0	0	0	27	0
8	0	0	9	0	0	0	0	27	0	0	0	36	9
9	9	0	27	0	0	0	0	9	9	0	9	63	54
10	0	9	9	0	0	0	0	27	0	108	0	153	45
total	108	36	162	54	54	27	72	108	36	108	9	774	297
comm	9	27	54	9	27	9	45	81	27	0	9	297	

(b) TM Data + Topography
(Elevation, Slope, and Incidence)

number pixels grouped into class													
class	1	2	3	4	5	6	7	8	9	10	uncl	tot	omis
1	117	0	0	0	0	0	0	0	0	0	0	117	0
2	0	9	0	0	9	0	0	0	0	0	0	18	9
3	0	0	144	0	9	0	0	[illegible]	0	0	9	171	27
4	0	0	0	81	0	0	0	0	0	0	0	81	0
5	0	0	18	0	36	0	9	0	0	0	0	63	27
6	0	0	0	0	0	36	9	0	0	0	0	45	9
7	0	0	0	0	0	0	27	0	0	0	0	27	0
8	0	0	0	0	0	0	0	36	0	0	0	36	0
9	0	0	27	0	0	0	0	9	18	0	9	63	45
10	18	9	18	0	0	0	0	0	0	108	0	153	45
total	135	18	207	81	54	36	45	54	18	108	18	774	162
Comm	18	9	63	0	18	0	18	18	0	0	18	162	

More poorly defined classes are the alpine meadow, alpine shrub, and mountain top classes. These classes primarily occur in areas of varying elevation, slope, and incidence angles. These variations ~~may~~ cause areas to look spectrally similar when in fact they represent different landscape classes. The converse may also be true; areas that represent the same landscape classes may appear spectrally different and consequently will be classified differently based on the sensor data.

Class 9 is very poorly defined based on the TM data alone. Only 14% of the class 9 pixels were classified correctly. Commission errors were also large with 27 of the 36 pixels identified as class 9 from the field data classified as other classes by the maximum likelihood classifier. This may be in part a reflection of the nature of this class. It occurs primarily in small patches across the study area, consequently it was difficult to develop training areas that adequately described the variability in this class. Further, deciduous vegetation often is located in valley bottoms where water and rocks are interspersed with the deciduous shrub and contribute to the response recorded by Landsat for pixels that represent this class.

The classification based on the integrated TM and topographic data set composed of elevation, slope, and incidence, has an overall accuracy of 77.1%; an increase of greater than 20% over the classification based on the Landsat sensor data alone. Improvements in individual class accuracies range from 14.1% in class 5 to 44.0% in class 4. A large reduction in omission errors for class 1 can be largely attributed to the extra discriminatory information provided by the slope variable. Pixels that are situated on steeper slopes on the alluvial plain tend to be better drained and consequently represent more mature spruce stands. In flat areas, drainage is often poor and organic material is more common. As a result, 18 pixels incorrectly identified as organic material on the TM classification, were identified correctly as mature spruce forest on the integrated map. In spite of this reduction in commission errors, class 2 remains poorly defined with only a 50% class accuracy. This is in part because limited training data were available to develop class signatures for this class and only 2 field sites (18 pixels) belonged to the class for testing class accuracy.

Class 3 improves by 21.0% when the topographic data set is employed in the classification process. A reduction in confusion between this class and class 9, which is spectrally similar but occurs at lower elevations in valley bottoms of the study area, contributes largely to this improvement.

Poor mapping accuracies for classes 5 and 9 after the topographic variables are incorporated may be attributed to several factors: (i) sites may have been incorrectly identified in the field or (ii) other information may be necessary to fully discriminate the landscape classes, for example, soil type, relief, or convexity measures. Since, however, in the preceding discriminant classification (see Chapter 5), classification accuracies were above 95% for these two classes, when topography was used, it is more likely that training areas for these classes may not have represented the full variance in sensor and topographic variables.

6.4. Summary

In this chapter, two separate maximum likelihood classifications were performed for the Yukon study area. The first was based on TM data alone; the second employed TM data combined with topographic data in the form of elevation, slope, and incidence extracted from a DEM. Maps were produced which show the spatial variations in the classification results. Mapping accuracy was determined by comparing the field classification with each maximum likelihood classification for the 774 pixels visited during the field season. For the classification base on TM data alone, mapping accuracy was 55.8%. This improved to 77.6% when the topographic data was added.

Results of this chapter support the results of Chapter 5 in providing evidence that topographic data is necessary for terrain classification of a high relief environment. In addition, these results show that the necessary topographic data can be extracted relatively easily from a DEM of the area since in the maximum likelihood analyses DEM data rather than topographic data derived from field measurements were employed. This is important from an operational viewpoint since in order to classify an entire area, continuous topographic coverage is required. In the future, it is possible that the continuous DEM will be available from the imagery themselves.

Chapter 7
Summary, Conclusions, and Recommendations

7.1. Summary

Landsat satellite imagery contains a topographic effect which, in mountainous terrain, can lead to poor classification accuracy. Attempts to reduce or eliminate this effect based on modelling have met partial success, particularly with MSS data, but TM data have improved spectral capabilities. Complex models designed to correct these data have not been adequate. Another idea is to integrate spectral response and topographic information from a DEM. This has been accomplished successfully for MSS data by Satterwhite et al. (1984), and others. More recently, Franklin (1987) based the integration of MSS and DEM data on the idea of integrated or landscape classifications first applied by Hutchinson (1978) and Robinove (1979) using satellite imagery.

The research discussed here was designed to address some of the methodological problems associated with TM data in the classification of high relief terrain. The main objective was to show that a data set composed of topographic terrain descriptors can provide additional information and lead to improved classification results if integrated with spectral data acquired by Landsat in terrain analysis of a high relief region.

The analysis is described in three stages. The first, correlation analysis, is used to document the relationships between the TM variables and the ground variables which include elevation, slope, incidence, and relief. Significant correlations indicated direct relationships between the various components of the landscape and the data recorded by Landsat. For example, the correlation between elevation and Band 5 was found to be r=0.58; coniferous vegetation correlates with Band 5 with r=-0.40. The fact that correlations were weak was interpreted to mean that the DEM variables might be expected to add discriminatory information in terrain classification. Little or no improvement in correlation results when both surface cover and topographic variables were considered as a group and examined with reference to the TM bands. This suggests that variance contained in the topographic data set that is contained in each TM band does not increase by the addition of the surface cover variables. This can be interpreted as the surface cover variables providing little information that is not already contained in the topographic data set and a single TM band.

The more complex canonical correlation analysis showed the extent to which the landscape classes occupied the same positions in the TM feature space as they did in the topographic and/or surface cover feature spaces. Since canonical correlations were significant between the TM and the surface cover (R_c=0.68 for the first vector pair) and the TM and the topographic data set (R_c=0.80), this suggested that there was similarity in the structure of landscape classes as defined using the TM and either of the other data sets. Consequently, integration of the TM and topographic data sets might provide a meaningful set of data useful for landscape classification. Since the canonical correlations are generally weak and moderate at best, and are not one to one, it is expected that additional information is contained in the topographic data set above that which is contained in the TM data set alone. This additional information may result in an improvement in the ability to discriminate between landscape classes when the integrated data set is employed.

The second stage, discriminant analysis, was designed to identify the actual improvements that are possible when the integrated data set was used. The strength of the statistical interpretation was examined by performing terrain classification based on various combinations of the TM, topographic and surface cover data sets. The addition of various topographic variables to the TM data set reduced errors of omission and commission and improved overall classification results over that which was obtained when the TM data were employed alone. When TM was used alone, classification accuracy was 66.5% for the training sample and 64.1% for the test sample. The best results were achieved when all topographic variables were incorporated. Overall classification accuracy improved to 97.5% and 98.7% for the training and test samples respectively. This represents an increase in classification accuracy of greater than 30 percent. The high level of accuracies achieved when the integrated data set was used illustrates that digital TM and topographic data can be used to improve terrain analysis in high relief environments.

A further demonstration of the critical role of the ancillary topographic data if terrain analysis is to be employed successfully in mountainous terrain was based on a spectrally consistent classification of a large area. The maximum likelihood classifications further support the important role of topography by demonstrating improvements in mapping accuracy for the study area when topographic data extracted from a DEM are integrated with the TM data. When the TM data were employed alone, overall mapping accuracy was only 55.8% based on 774 pixels studied in the field. This increased to 77.6% when elevation, slope, and incidence variables were included. The largest improvement was 46% for class 4 (Alpine Meadow) with more than 10% improvement in each of the other classes except class 10 (Immature Spruce) which remained the same.

7.2. Conclusions

Results of the analysis reported in this thesis support the following conclusions:

(i) There are significant relationships between surface cover and topography for the area selected for this research. Correlation coefficients were interpreted as evidence that an analysis approach that considers both topography and surface cover characteristics, a landscape or biophysical approach, is appropriate for classification of this high relief environment.

(ii) Relationships between ground variables and sensor variables are weak to moderate. These relationships were interpreted to mean that the ground variables contain variance that is unique compared to the sensor data and consequently, these variables might provide additional information if incorporated in terrain analysis.

(iii) Canonical correlation coefficients indicated that common variance between the ground and sensor data sets are relatively small; 46% for surface cover and 64% for topography. The conclusion based on these correlations was that the sensor data alone do not contain sufficient information to discriminate between landscape classes defined according to the integrated or landscape approach. Consequently, these results show that additional information is required if terrain analysis is to be carried out successfully in this high relief region. Since continuous topographic data can be obtained relatively easily from a DEM, an ancillary topographic data set is the more practical source of additional information.

(iv) Little improvement in correlation coefficients was found when surface cover variables were used in addition to the topographic variables in the canonical model. This was interpreted as evidence that much of the variance in surface cover was explained by the topographic and sensor variables alone. Further, it was concluded that an ancillary topographic data set should be incorporated in the analysis of high relief terrain for improved classification results.

(v) Landscape classification of the Southwest Yukon study area using Landsat TM data alone resulted in maps of low accuracy, only 66.5% at best based on an analysis of a random set of 672 training pixels for which the associated landscape classes were known from the field and identified on the digital classification. Accuracy was 64.1% based on an independent test data set.

(vi) Use of an integrated TM and topographic data set for discriminating among landscape classes in this high relief environment improved classification results significantly over when the TM data were employed alone. Overall classification results improved by more than 30% with the addition of elevation, slope, incidence, and relief variables.

(vii) The optimal topographic data set, or that which produced the best classification results based on the known field data, incorporated all the topographic variables examined in this analysis, ie. elevation, slope, incidence, and relief. Results for the training data set were 97.5%. The corresponding accuracy when either topographic variable was added alone was at most 77.7%.

(viii) Mapping accuracy improved significantly for the Yukon study area when a topographic data set extracted from a DEM was incorporated in the classification procedure over when the TM data were used alone. Using the TM data set alone resulted in a mapping accuracy of only 55.8% based on 774 pixels visited in the field. This improved to 77.6% with the addition of elevation, slope, and incidence.

(ix) Based on the overall results of this research, it can be concluded that an integrated data set composed of topographic and sensor data can be employed to improve classification results in high relief terrain analysis.

7.3. Recommendations for Future Research

The research described in this thesis identified improvements in TM classification results with the addition of ancillary topographic information for an area of high relief terrain. Another area of research might involve the development of improved methods of training area selection for supervised classification. For example, methods which reduce subjective analyst interpretations and ensure that the variability within terrain classes of interest is captured for each variable. This is particularly important when integrated data sets are employed and numerous terrain characteristics considered in the selection of training areas. One possible suggestion to reduce operator subjectivity might be to convert sensor data to absolute values such as albedo. If known albedo values for various land cover types could be established, then the operator would not be responsible for selecting training areas based on his/her own knowledge of the region. Rather, known albedo values could be used to develop class signatures. This is fine for the sensor data but limits on the ancillary data variables would also have to be established when integrated data sets are used.

The ability to mask out certain sections in the study area when the integrated data sets are used may improve discrimination of classes where certain terrain descriptors are not important or add confusion to the classifiers. In this analysis, for example, the addition of topographic variables beyond elevation, resulted in lower classification accuracies than when the sensor variables and elevation alone were employed to map class 1 (Forest Plain). If classifiers could be developed to incorporate certain variables for the discrimination of some classes and different combinations for the discrimination of others, overall classification results may be improved.

Research is required into the question of relationships between topographic data recorded or derived from measurements taken in the field and similar data extracted from a DEM. Various methods of producing DEMs should be examined to determine which produces results which correlate best with the field data. In this area, alternative sources of topographic data (other than topographic maps) for example, DEMs produced using photographic correlation machines or from stereo SPOT satellite data should be considered. Different interpolation routines for creating elevation grids from digitized contours can be examined and/or developed; and finally, other topographic variables, for example, convexity should be examined.

Finally, the research described in this thesis examined one approach to terrain classification; mapping landscape or integrated terrain units using Landsat Thematic Mapper and Digital Elevation Model Data. Further research is needed in other applications such as classifying and mapping terrain classes described in terms of glaciated units or defined from a geomorphological or geological perspective. Future research might also consider different types of satellite data, for example, SPOT.

References

Ahern, F.J. (1985). Preprocessing for Multi-source Data Integration. *Ninteenth International Symposium on Remote Sensing of Environment, Proceedings.* Ann Arbor, Michigan, 95-103.

Ahern, F.J., Brown, R.J., Cihlar, J., Gauthier, R. Murphy, J., Neville, R.A., and Teillet, P.M. (1987). Radiometric Correction of Visible and Infrared Remote Sensing Data at the Canada Center for Remote Sensing. *International Journal of Remote Sensing, 8,* 1349-1376.

Bastedo, J.D. and Theberge, J.B. (1983). An Appraisal of Inter-Disciplinary Resource Surveys - Ecological Land Classification. *Landscape Planning, 10,* 317-334.

Bastedo, J.D., Nelson, J.G. and Theberge, J.B. (1984). Ecological Approach to Resource Survey and Planning for Environmentally Significant Areas : The ABC Method. *Environmental Management, 8*(2), 125-134.

Bernstein, R. (Ed). (1978). *Digital Image Processing for Remote Sensing.* New York: Institute of Electrical and Electronics Engineers, Inc. 473 pp.

Bonner, W.J., Rohde, W.G., and Miller, W.A. (1982). Mapping Wildland Resources with Digital Landsat and Terrain Data. In *Remote Sensing and Resource Management.* Iowa: SCSA. 73-80.

Brinker, R.C. and Wolf, P.R. (1974). *Elementary Surveying.* New York: Harper and Row. 608pp.

Burrough, P.A. (1986). *Principles of Geographic Information Systems for Land Resource Assessment Monographs on Soil and Resource Survey No.12.* Oxford: Clarendon Press. 193.

Cavayas, Francois. (1987). Modelling and Correction of Topographic Effect Using Multitemporal Satellite Images. *Canadian Journal of Remote Sensing, 13*(2), 49-67.

Chavez, P.S.Jr., Everett, J.R., Kahle, A.B., Kieffer, H.H. Kitcho, C.A., Milton, N.M. and Mouat, D.A. (1983). Thematic Mapper Data Analysis. In *Frontiers for Geological Remote Sensing from Space, Report of the Fourth Geosat Workshop.* Flagstaff, Arizona: American Society of Photogrammetry. 21-26.

Christian, C.S. (1958). The Concept of Land Units and Land Systems. *Ninth Pacific Science Congress, Proceedings, 20,* 74-81.

Christian, C.S. and Stewart, G.A. (1968). Methodology of Integrated Surveys. *Aerial Surveys and Integrated Studies, Proceedings.* Toulouse, France, 233-280.

Cibula, W.G. and Nyquist, M.O. (1987). Use of Topographic and Climatological Models in a Geographical Data Base to Improve Landsat MSS Classification of Olympia National Park. *Photogrammetric Engineering and Remote Sensing, 53*(1), 67-75.

Clarke, D. (1975). *Understanding Canonical Correlation Analysis: Concepts and Techniques in Modern Geography No.3.* Norwich: GeoAbstracts. 36pp.

Collins, S.H. (1975). Terrain Parameters Directly from a Digital Terrain Model. *Canadian Surveyor, 9*(5), 507-518.

Collins, S.H. and Moon, G.C. (1981). Algorithms for Dense Grid Digital Terrain Models. *Photogrammetyric Engineering and Remote Sensing, 47*(1), 71-76.

Colwell, R.N. (Ed) (1983). *Manual of Remote Sensing Second Edition.* Falls Church, Virginia: American Society of Photogrammetry. 2440 pp.

Conner, P.K. and Mooneyhan, D.W. (1985). Practical Applications of Landsat Data. In *Monitoring Earth's Ocean, Land, and Atmosphere from Space - Sensors, Systems, and Applications.* New York: American Institute of Aeronautics and Astronautics, Inc. 371-396.

Cooper, P.R., Friedman, D.E. and Wood, S.A. (1985). *The Automatic Generation of Digital Terrain Models from Satellite Images by Stereo.* Richmond, British Columbia: MacDonald Dettwiler and Associates. 10 pp.

Crawley, B.G. (1974). Gestalt Contours. *Canadian Surveyor, 28*(3), 237-246.

Curran, P.J. (1985). *Principles of Remote Sensing.* New York: Longman Inc. 282pp.

Curran, O.J. and Williamson, H.D. (1085). The Accuracy of Ground Data Used in Remote Sensing Investigations. *International Journal of Remote Sensing, 6*, 1637-1651.

Dave, J.V. and Bernstein, R. (1982). Effect of Terrain Orientation and Solar Position on Satellite-Level Luminance Observations. *Remote Sensing of Environment, 12*, 331-348.

Deane, G.C., Churchill, P.N., and Griffiths, G.H. (1985). Preprocessing for Multi-source Data Integration. *Ninteenth International Symposium on Remote Sensing of Environment, Proceedings.* Ann Arbor, Michigan, 633-641.

DIPIX. (1987). *ARIES System User's Manual.* Ottawa, Ontario: DIPIX Systems Limited. 3 Vols.

Dottavio, C.L. (1981). Preprocessing for Multi-source Data Integration. *Seventh Symposium on Machine Processing of Remotely Sensed Data, Proceedings*. LARS, Purdue University, 375-383.

Engel, J.L. and Weinstein, O. (1983). Thematic Mapper - An Overview. *IEEE Transactions on Geoscience and Remote Sensing*, *GE*-21(3), 258-265.

Evans, I.S. (1972). General Geomorphometry, Derivatives of Altitude and Descriptive Statistics. In *Spatial Analysis in Geomorphology*. London: Methuen. 17-90.

Fleming, M.D. and Hoffer, R.M. (1979). Machine Processing of Landsat MSS and DMA Topographic Data for Forest Cover Type Mapping. *Fifth Symposium on Machine Processing of Remotely Sensed Data, Proceedings*. West Lafayette, Indiana, 377-390.

Franklin S.E. and LeDrew E.F. (1984a). An Assessment of Information from DEM Variables and Landsat MSS Data for an Area of High Relief. *Ninth Canadian Symposium on Remote Sensing, Proceedings* . St.John's, Newfoundland, 451-460.

Franklin S.E. and LeDrew E.F. (1984b). High Relief Terrain Classification Using Digital Elevation Model Variables and Landsat MSS Data in the Yukon Territory, Canada. *Eighteenth International Symposium on Remote Sensing of Environment, Proceedings* . Paris, France, 603-611.

Franklin, S.E. (1987). Terrain Analysis from Digital patterns in Geomorphometry and Landsat MSS Spectral Response. *Photogrammetric Engineerong and Remote Sensing, 53*(1), 59-65.

Franklin, S.E. and Peddle, D.R. (1987). Texture Analysis of Digital Image Data Using Spatial Cooccurrence. *Computers and Geosciences, 13*(3), 293-311.

Franklin S.E., Rogerson, R.J., and Moulton, J.E. (1987). Interpretation of High Relief Glaciated Environments Using Digital Spectral and Geomorphometric Data. *Eleventh Canadian Symposium on Remote Sensing, (article accepted for presentation).* Waterloo, Ontario, .

Franklin, S.E., Peddle, D.R, and Moulton, J.E. (1988). Spectral/Geomorphometric Discrimination and Mapping of Terrain : A Study in Gros Morne National Park. *Canadian Journal of Remote Sensing,* (), . (in press).

Freden, S.C. and Gordon, F. (1983). Landsat Satellites. In *Manual of Remote Sensing.* Falls Church, Virginia: American Society of Photogrammetry. 517-570.

Gregory, A.F. and Moore, H.D. (1986). Thematic Mapping from Landsat and Collateral Data : A Review of one Company's Experience and a Forecast of Future Potential. *Canadian Journal of Remote Sensing, 12*(1), 55-63.

Hall-Konyves, Karin. (1987). The Topographic Effect on Landsat Data in Gently Undulating Terrain in Southern Sweden. *International Journal of Remote Sensing, 8*(2), 157-168.

Hammond, R. and McCullagh, P.S. (1980). *Quantitative Techniques in Geography - An Introduction.* Oxford: Clarendon Press. 362 pp.

Haralick, R.M., Shanmugam, K., and Dinsein, Its'hak. (1973). Textural Features for Image Classification. *IEEE Transactions on Systems, Man, and Cybernetics, SMC-3*(6), 610-621.

Helwig, J.T. and Kathryn, A.C. (Ed). (1979). *SAS User's Guide*. Cary, North Carolina: SAS Institute Inc. 494 pp.

Hinse, M., Gwyn, Q.H., and Bonn, F. (1988). Radiometric Correction of C-Band Imagery for Topographic Effects in Regions of Moderate Relief. *IEEE Transactions on Geoscience and Remote Sensing, GE-26*, 122-132.

Holben, B.N. and Justice, C.O. (1980). The Topographic Effect on Spectral Response from Nadir-Pointing Sensors. *Photogrammetric Engineering and Remote Sensing, 46*(9), 1191-1200.

Holben, B.N. and Justice, C.O. (1981). An Examination of Spectral Band Ratioing to Reduce the Topographic Effect on Remotely Sensed Data. *International Journal of Remote Sensing, 2*(2), 115-133.

Horler, D.H. and Ahern, F.J. (1986). Forestry Information Content of Thematic Mapper Data. *International Journal of Remote Sensing, 7*, 405-428.

Hutchinson, C.F. (1978). *The Digital Use of Landsat Data for Integrated Land Resource Survey : A Study in the Eastern Mojave Desert, California.* Doctoral dissertation, University of California, Riverside, 265 pp.

Hutchinson, C.F. (1982). Techniques for Combining Landsat Data and Ancillary Data for Digital Classification Improvement. *Photogrammetric Engineering and Remote Sensing, 48*(1), 123-130.

Jones, A.R., Settle, J.J., and Wyatt, B.K. (1988). Use of Digital Terrain Data in the Interpretation of SPOT-1 HRV Multispectral Imagery. *International Journal of Remote Sensing, 9*, 669-682.

Justice, C.O. (1978). An Examination of the Relationship Between Selected Ground Properties and Landsat MSS Data in an area of Complex Terrain in

Southern Italy. *American Society of Photogrammetry, Fall Technical Meeting, Proceedings*. Albuquerque, New Mexico, 303-328.

Justice, C.O., Wharton, S.W., and Holben, B.N. (1981). Application of Digital Terrain Data to Quantify and Reduce the Topographic Effect on Landsat Data. *International Journal of Remote Sensing*, *2*(3), 213-230.

Justice, C.O. and Townshend, J.R.G. (1981). Integrating Ground Data with Remote Sensing. In *Terrain Analysis and Remote Sensing*. London: George Allen and Unwin Limited. 38-58.

Kan, E.D. and Weber, F.P. (1978). The Ten Ecosystem Study : Landsat ADP Mapping of Forest and Rangeland in the United States. *Twelfth International Symposium on Remote Sensing of Environment, Proceedings*. Ann Arbor, Michigan, 1809-1826.

Karaska, M.A., Walsh, S.J., and Butler, D.R. (1986). Impact of Environmental Variables on Spectral Signatures Acquired by the Landsat Thematic Mapper. *International Journal of Remote Sensing*, *7*(12), 1653-1667.

Kawata, Y., Ueno, S., and Kusaka, T. (1985). Removal of Atmospheric and Topographic Effects from Landsat MSS Imagery. *Ninteenth International Symposium on Remote Sensing of the Environment, Proceedings*. Ann Arbor, Michigan, 821-828.

Khorram, Siamak, Brockhaus, J.A., and Chesire, H.M. (1987). Comparison of Landsat MSS and TM data for Urban Land-Use Classification. *IEEE Transactions on Geoscience and Remote Sensing*, *GE-25*, 238-243.

Klecka, W.R. (1980). *Discriminant Analysis*. Beverly Hills and London: Sage Publications. 71pp.

Mabbutt, J.A. (1968). Review of Concepts of Land Classification. In *Land Evaluation - Papers of a CSIRO Symposium organized in cooperation with UNESCO 26-31.* Melbourne: MacMillan of Australia. 11-28.

Mayer, K.E., Fox, L. III, and Webster, J. (1979). Forest Condition Mapping of the Hoopa Valley Indian Reservation Using Landsat Data. *First International Symposium of Remote Sensing for Natural Resources, Proceedings.* University of Idaho, 217-242.

Moulton, J.E. (1986). *Use of Calibrated Landsat Data for Improving the Accuracy of Subarctic Terrain Classification, Unpublished Thesis.* St.John's, Newfoundland: Memorial University of Newfoundland. 46 pp.

Peddle, D.R. (1987). *The Digital Integration of Topographic Information and Spectral Response Patterns for Terrain Classification, Unpublished Thesis.* St.John's, Newfoundland: Memorial University of Newfoundland. 94 pp.

Richards, J.A. (1986). *Remote Sensing Digital Image Analysis.* Berlin: Springer-Verlag. 281 pp.

Robinove, C.J. (1979). *Integrated Terrain Mapping with Digital Landsat Images in Queensland, Australia - Geological Survey Professional Paper 1102.* Washington: United States Government Printing Office. 39 pp.

Robinove, C.J. (1981). The Logic of Multispectral Classification and Mapping of Land. *Remote Sensing of Environment, 11*, 231-244.

Robinove, C.J. (1982). Computation with Physical Values from Landsat Digital Data. *Photogrammetric Engineering and Remote Sensing, 48*(5), 781-784.

Sampson, R.J. (1978). *Surface II Graphics System.* Lawrence, Kansas: Kansas Geological Survey. 240 pp.

SAS Institute Inc. (1985). *SAS User's Guide : Basics Version 5 Edition.* Cary, NC: SAS Institute Inc. 1290pp.

Satterwhite, M., Rice, W., and Shipman, J. (1984). Using Landform and Vegetation Factors to Improve the Interpretation of Landsat Imagery. *Photogrammetric Engineering and Remote Sensing, 50*(1), 83-91.

Seidel, K., Ade, F., and Lichtenegger, J. (1982). Augmenting Landsat-MSS Data with Topographic Information for Enhanced Registration and Classification. *International Geoscience and Remote Sensing Symposium, IGARSS'82, Proceedings.* Munich, West Germany, 2.1-2.6.

Shasby, Mark and Carneggie, David. (1986). Vegetation and Terrain Mapping in Alaska Using Landsat MSS and Digital Terrain Data. *Photogrammetric Engineering and Remote Sensing, 52*(6), 779-786.

Smith, J.A., Tzeu, Lie Lin, and Ranson, K.J. (1980). The Lambertian Assumption and Landsat Data. *Photogrammetric Engineering and Remote Sensing, 46*(9), 1183-1189.

Stohr, C.J. and West, T.R. (1985). Terrain and Look Angle Effects Upon Multispectral Scanner Response. *Photogrammetric Engineering and Remote Sensing, 51*(2), 229-235.

Strahler, A.H., Logan, T.L., and Bryant, N.A. (1978). Improving Forest Cover Classification Accuracy from Landsat by Incorporating Topographic Information. *Twelfth International Symposium on Remote Sensing of the Environment, Proceedings.* Ann Arbor, Michigan, 927-956.

Strahler, A.H., Estes, J.E., Maynard, P.F., Mertze, F.C., and Stow, D.A. (1980). Incorporating Collateral Data in Landsat Classification and Modelling Procedures. *Fourteenth International Symposium on Remote Sensing of Environment, Proceedings*. An Arbor, Michigan, 1009-1026.

Swain, P.H. and Davis, S.M. (Ed). (1978). *Remote Sensing : The Quantitative Approach*. New York: McGraw-Hill Inc. 396 pp.

Teillet, P.M., Guindon, B., and Goodenough, D.G. (1982). On the Slope-Aspect Correction of Multispectral Scanner Data. *Canadian Journal of Remote Sensing, 8*(2), 84-106.

Theberge, J.B. (1980). *Kluane : Pinnacle of the Yukon*. Toronto: Doubleday Canada Limited. 175 pp.

Thorndike, R.M. (1978). *Correlation Procedures for Research*. New York: Gardner Press, Inc. 340pp.

Toll, D.L. (1985). Effect of Landsat Thematic Mapper Sensor Parameters on Landcover Classification. *Remote Sensing of Environment, 17*(2), 129-140.

Tom, C.H. and Miller, L.D. (1984). An Automated Land-use Mapping Comparison of the Bayesian Maximum Likelihood and Linear Discriminant Analysis Algorithms. *Photogrammetric Engineering and Remote Sensing, 50*(2), 193-207.

Townshend, J.R.G. and Justice, Chris. (1981). Information Extraction from Remotely- Sensed Data, A User View. *International Journal of Remote Sensing, 2*(4), 313-329.

Walsh, S. (1987). Variability of Landsat MSS Spectral Responses of Forests in Relation to Stand and Site Characteristics. *International Journal of Remote Sensing, 8*(9), 1289-1299.

Wang, Shuyan, Haralick, R.M., and Cambell, J.B. (1984). Relative Elevation Determination from Landsat Imagery. *Photogrammetria, 39*(4-6), 193-215.

Webster, R. and Beckett, P.H.T. (1970). Terrain Classification and Evaluation Using Air Photography : A Review of Recent Work at Oxford. *Photogrammetria, 26*(2/3), 51-71.

Woodcock, C.E. (1980). Stratification of Forest Vegetation for Timber Inventory Using Landsat and Collateral Data. *Fourteenth International Symposium on Remote Sensing of Environment, Proceedings.* San Jose, Costa Rica, 1769-1787.

Woodcock, C.E. (1982). *Reducing the Influence of Topography on the Classification of Remotely Sensed Data, M.A. Thesis, Unpubl.*. Santa Barbara, California: University of California, Santa Barbara. 69 pp.

Appendix A
Programs

```
C    PROGRAM NAME:          OUTWDW.FOR
C    Author:                Joan E. Moulton
C                           Department of Geography
C                           Memorial University of Newfoundland
C
C    Description:           This program is designed to output TM pixel
C                           values of a 3 x 3 window given the
C                           coordinates of the center pixel.
C
C    Computer:              VAX-8800 and VAX-11/785
C    Language:              Fortran 77 (VAX)
C    Operating System:      VMS (version 4.5)
C    Date:                  14 Nov 1987
C    Execution Sequence:
C                           $ fortran outwdw.for
C                           $ link outwdw
C                           $ run outwdw

      integer xco(99), yco(99), i, x, y, l, j, k, xm, xp, ym, yp
      byte    barr(551,548)
      character*15 infile
      character*15 outfile

      print*,'Enter input band file'
      accept 60, infile
      print*,'Enter outfile'
      accept 60, outfile

      open(unit=1, file='coord.dat', status='old')
      open(unit=2, file=infile, status='old')
      open(unit=3, file=outfile, status='new')

      do 10 y=1,551
         read(2,200)  (barr(y,x),x=1,548)
```

```
10        continue
          do 20 j=1,100
             read(1,*) xco(j), yco(j)
                     x = xco(j)-374
                     y = yco(j)-849
                    xp = x+1
                    xm = x-1
                    yp = y+1
                    ym = y-1
          do 30 l = ym, yp
              do 40 k = xm, xp
                    if ((x.ne.-1).and.(y.ne.-1)) then
                           i = barr(l,k)
                           i=iand(i,255)
                           write(3,*) i
                  else
                  write(3,*) x
                  endif
40      continue
30      continue
20      continue
200     format(548a1)
60      format(a15)
           close(unit=1)
           close(unit=2)
           close(unit=3)
           stop
           end
```

```
C     PROGRAM NAME:          SLOPE.FOR
C     Author:                Joan E. Moulton
C                            Department of Geography
C                            Memorial University of Newfoundland
C
C     Description:           This program is designed to calculate an
C                            average slope plane given the elevations
C                            in a 3X3 pixel window.
C
C     Computer:              VAX-8800 and VAX-11/785
C     Language:              Fortran 77 (VAX)
C     Operating System:      VMS (version 4.5)
C     Date:                  11 Dec 1987
C     Execution Sequence:
C                            $ fortran slope.for
C                            $ link slope
C                            $ run slope

      integer n1, n2, n3, n4, n5, n6, n7, n8, n9
      integer i, r, slope
      real temp1, temp2, temp3, temp4, sl

      open(unit=1, file='elevarr.dat', status='old')
      open(unit=2, file='slope.dat', status='new')

      do 10 i=1,100
       read(1,*) n1, n2, n3, n4, n5, n6, n7, n8, n9
      if (n1 .ge. 0)  then
        temp1=(n1-n7)+(n2-n8)+(n3-n9)
        temp2=(n1-n3)+(n4-n6)+(n7-n9)
        temp3=(temp1/180)*(temp1/180)+(temp2/180)*(temp2/180)
        temp4=sqrt(temp3)
        sl=atand(temp4)
        slope=int(sl)
              else
               slope=-1
              endif
        do 40 r=1,9
        write(2,*) slope
40      continue
10      continue
        close(unit=1)
        close(unit=2)
        stop
        end
```

```
C  PROGRAM NAME:            ASPECT.FOR
C  Author:                  Joan E. Moulton
C                           Department of Geography
C                           Memorial University of Newfoundland
C
C  Description:             This program is designed to calculate the
C                           average direction that the slope plane
C                           caclulated using SLOPE.FOR faces.
C
C  Computer:                VAX-8800 and VAX-11/785
C  Language:                Fortran 77 (VAX)
C  Operating System:        VMS (version 4.5)
C  Date:                    19 Jan 1988
C  Execution Sequence:
C                           $ fortran aspect.for
C                           $ link aspect
C                           $ run aspect

      integer n1, n2, n3, n4, n5, n6, n7, n8, n9
      integer i, r, aspect
      real temp1, temp2, temp3, temp4, temp5, asp

      open(unit=1, file='elevarr.dat', status='old')
      open(unit=2, file='aspect.dat', status='new')

      do 10 i=1,100
        read(1,'(9 i5)') n1, n2, n3, n4, n5, n6, n7, n8, n9

       if (n1 .ge. 0) then
        temp1=((n1-n7)+(n2-n8)+(n3-n9))
        temp2=((n1-n3)+(n4-n6)+(n7-n9))
        temp3=-1*temp1/180
        temp4=-1*temp2/180

          if (temp3 .eq. 0) then
            if (temp4 .eq. 0) then
            aspect=0
            else
             if (temp4 .gt. 0) then
             aspect=0
             else
             aspect=180
             endif
            endif
          else
```

```
                temp5=temp4/temp3
                asp=atand(temp5)
                aspect=int(asp)
                if (aspect .ge. 0) then
                  if ((temp3 .lt. 0) .and. (temp4 .le. 0)) then
                  aspect=aspect+90
                  else
                  aspect=aspect+270
                  endif
                else
                    if (temp3 .gt. 0) then
                    aspect=aspect*-1+180
                    else
                    aspect=aspect*-1
                    endif
                endif
          endif
        else
        aspect=999
        endif
         print*, aspect
         do 20 r=1,9
         write(2,*) aspect
20       continue
10       continue
         close(unit=1)
         close(unit=2)
         stop
         end
```

```
C    PROGRAM NAME:        INC.FOR
C    Author:              Joan E. Moulton
C                         Department of Geography
C                         Memorial University of Newfoundland
C
C    Description:         This program is designed to calculate
C                         incidence values as a function of slope,
C                         aspect, sun elevation and sun angle at the
C                         time of sensor overpass.
C
C    Computer:            VAX-8800 and VAX-11/785
C    Language:            Fortran 77 (VAX)
C    Operating System:    VMS (version 4.5)
C    Date:                20 Jan 1988
C    Execution Sequence:
C                         $ fortran inc.for
C                         $ link inc
C                         $ run inc

      integer aspect, i, j, incidence
      real sunelev, azidiff, slope, azimuth, inc

      open(unit=1, file='slope.dat', status='old')
      open(unit=2, file='aspect.dat', status='old')
      open(unit=3, file='inc.dat', status='new')
      sunelev=42.0
      azimuth=154.0
      do 100 j=1,900
         read(1,*) slope
         read(2,*) aspect

         if ((slope .ne. -1) .and. (aspect .ne. 999)) then
         azidiff=aspect-azimuth
         inc=cosd(slope)+sind(slope)*cosd(sunelev)*cosd(azidiff)
         incidence=int(inc)
        else
         inc=-1
         endif
         write(3,'(f7.3)') inc
100      continue
       close(unit=1)
       close(unit=2)
       close(unit=3)
       stop
       end
```

```
C     PROGRAM NAME:          RELIEF.FOR
C     Author:                Joan E. Moulton
C                            Department of Geography
C                            Memorial University of Newfoundland
C
C     Description:           This program is designed to calculate
C                            relief values for each pixel based on the
C                            variance in elevation at each site.
C
C     Computer:              VAX-8800 and VAX-11/785
C     Language:              Fortran 77 (VAX)
C     Operating System:      VMS (version 4.5)
C     Date:                  20 Jan 1988
C     Execution Sequence:
C                            $ fortran relief.for
C                            $ link relief
C                            $ run relief

          integer n1, n2, n3, n4, n5, n6, n7, n8, n9
          integer i, j, relief
          real mean, temp, ms, s

          open(unit=1, file='elevarr.dat', status='old')
          open(unit=2, file='relief.dat', status='new')

          do 10 i=1,100
            read(1,*) n1, n2, n3, n4, n5, n6, n7, n8, n9
            mean = (n1 + n2 + n3 + n4 + n5 + n6 + n7 + n8 + n9)/9
            ms = mean * mean
            temp=(n1*n1-ms)+(n2*n2-ms)+(n3*n3-ms)+(n4*n4-ms)+
     *       (n5*n5-ms)+(n6*n6-ms)+(n7*n7-ms)+(n8*n8-ms)+(n9*n9-ms)
            s=sqrt(temp/9)
            relief=int(s)
                      print*, relief
                  do 20 j=1,9
                      write(2,'(f6.2)') s
20          continue
10          continue
            close(unit=1)
            close(unit=2)
            stop
            end
```

Appendix B
Confidence Intervals

B.1. Explanatory Note to Tables 5-2 and 5-4

Example Calculation : TM Data Alone - Training Data

N = number of samples
P = number correct
Q = number of incorrect
p = probability of correctly classifying a pixel
q = probability of incorrectly classifying a pixel
m = mean for the binomial distribution
s = standard deviation for the binomial distribution
e_m = standard error of mean estimate
e_s = standard error of estimate of standard deviation

N = 672, P = 438, Q = 234

$p = P/N = .6518$
$q = Q/n = .3482$
$m = np = 438.010$
$s = SQR(npq) = 12.350$
$e_m = s/N = .001$
$e_s = s/SQR(2N) = .337$

Lower acceptable limit to give a 99.9% confidence level
$CL_{lower} = [(m - 3e_m) - 3(s + 3e_s)/N = 59.10\%$

Upper limit :
$CL_{upper} = [(m - 3e_m) + 3(s + 3e_s)]/N = 71.14\%$

Confidence Interval is : $CL_{lower} - CL_{upper}$ = 59.10% - 71.14%

Conclusion : We are 99.9% sure that the TM classification based on the training data is at least 59.10% accurate, but not more than 71.14% accurate when compared with the site data.]

Confidence Intervals for Selected Discriminant Classifications

Function	Confidence Interval Training Data	Test Data
(a) TM Alone	59.10% - 71.14%	45.22% - 79.98%
(b) Topography Alone	75.70% - 85.59%	61.10% - 93.42%
(g) TM + Elevation, Slope, and Incidence	85.41% - 93.16%	52.95% - 86.99%
(h) TM + All Topography	95.15% - 97.60%	93.02% - 100.00%

B.2. Explanatory Note to Tables 6-2 and 6-3

Confidence Intervals for Maximum Likelihood Classifications

Function	Confidence Interval
(a) TM Alone	56.00% - 67.27%
(b) TM + Elevation, Slope, and Incidence	74.34% - 83.79%